Supplementary Appendix (Not For Publication)

Omitted Proofs

Proof of Lemmas 1 and 3. We consider the model with one-sided commitment (Lemma 3). The proof of the two-sided commitment case (Lemma 1), which follows similar steps but is simpler, is omitted.

Suppose the period-t self of the consumer offers a contract $\mathcal{C}'_t$. Specifically, a contract at time t, $\mathcal{C}'_t$, specifies consumption on each possible state in each future time $\tau \geq t$. Denote the set of possible states by $K_{t,\tau}$, in which the first subscript corresponds to the time in which the contract is offered and the second subscript corresponds to the decision-making time τ. The contract specifies consumption for each different income states, so the contracting space is generally greater than the space of income states. In addition, perception-perfect equilibrium imposes no restrictions on $K_{t,\tau}$, i.e., $K_{t,\tau}$ can be arbitrary. To keep analysis tractable, we assume that $K_{t,\tau}$ has a product structure and only depends on decision making time τ. Otherwise, we can always add more states that are never reached so that it has a product structure and the resulting equilibrium is outcome-equivalent to the original equilibrium. Specifically, we write $K_{t,\tau} = \mathbb{S}_\tau \times H_\tau$, in which H_τ consists of all the possible income-independent messages/actions that the agent can send at time τ. The income-independent messages can be arbitrary. One of the reasons that an income-independent message can arise is from the consumer's different beliefs. Since we allow any contracts, we cannot impose what types of income-independent messages the consumer can send. For simplicity, we call H_τ the income-independent history. Without loss of generality, $H_1 = \emptyset$. Denote h_t a generic element in H_t. We call h_t an income-independent message. Denote $H_\tau(h_t)$ the states that can be reached at time τ from an earlier history $h_t \in H_t$ for $\tau > t$.

Fix a contract, we next write down the agent's strategy profile. Consider an agent who makes a decision at time τ. Suppose the income-independent messages that has been reached is $h_{\tau-1}$, which is an element in $H_{\tau-1}$. At time τ, the agent learns the income state, i.e., s_τ is realized. The agent needs to decide which message $a_\tau \in \Delta(H_\tau(h_{\tau-1}))$ to send, where $\Delta(\cdot)$ represents the set of lotteries. If there is one-sided commitment, the agent also needs to decide whether he will lapse or not, in which case, the strategy can be summarized by a pair (d_τ, a_τ), where $d_\tau \in \Delta(\{0,1\})$. If $d_\tau = 1$ with probability 1, then the agent stays, otherwise the contract is lapsed with a positive probability.

As described in the body of the paper, the perception-perfect equilibrium is solved by treating the agent's decisions in each period as if it were taken by a different player (i.e., a different "self"). The main claim is that for any perception-perfect equilibrium, the consumption vector must solve the program (P').

For the ease of exposition, we say that two perception-perfect equilibria are *equivalent* if all selves of the consumers have same actual and perceived consumption. We will establish the result through two separate claims:

Claim 2. *Fix a perception-perfect equilibrium. There exists an equivalent perception-perfect equilibrium in which the agent never lapses ($d_\tau = 1, \forall\tau$).*

Proof. Consider a perception-perfect equilibrium in which the agent lapses in some period $d_\tau = 0$ with a positive probability, replacing it with a contract $\mathcal{C}''_\tau$ from another firm. Since the other firm cannot lose money by offering this new contract, the old firm could have accepted a contract that substituted the terms of the old contract from this period on with the terms of the new contract, and the agent would have accepted to remain with the old firm. The constructed new contracts together with the agent's optimal decision forms a perception-perfect equilibrium that is equivalent to the original one. □

Claim 3. *Fix a perception-perfect equilibrium. There is an equivalent perception-perfect equilibrium that offers two options following any history: $\#|H_t(h_{t-1})| \le 2$, for all $h_{t-1} \in H_{t-1},\ t \ge 2$.*

Proof. From the previous claim, we can restrict attention to equilibria in which the agent never lapses. Suppose $t_1 < t_2 < t_3$. Note that self t_1's prediction about self t_3's decision coincides with self t_2's prediction about self t_3's decision. Restricting $H_t(h_{t-1})$ to two messages – one that the agent will choose and another one that the agent thinks that he will choose – does not affect the actual consumption or the perceived consumption. Put differently, if $H_t(h_{t-1})$ has at least three messages, then there is at least one of them that the agent never sends and the agent never believes other selves would send. Therefore, we can restrict the income-independent message space to be at most two: one that the agent actually choose, and one that the agent thought he would choose. □

Given these two claims, a contract offered by self t, $\mathcal{C}'_t$, must maximize the agent's utility subject to the zero profits, incentive compatibility, perceived choice, and non-lapsing constraints, concluding the proof of Lemma 3.

□

Proof of Lemmas 2 and 4. In the text, we presented the proof for the case with two-sided commitment when there is no uncertainty and $T = 4$. Here, we consider the model with one-sided commitment case (Lemma 4), still assuming no uncertainty and $T = 4$. The proof for stochastic income and arbitrary T is presented in the supplementary appendix.

There are two ICs:

$$u(c_2(A)) + \beta[\delta u(c_3(A,B)) + \delta^2 u(c_4(A,B))] \geq u(c_2(B)) + \beta[\delta u(c_3(B,B)) + \delta^2 u(c_4(B,B))], \tag{sA1}$$

$$u(c_3(A,A)) + \beta\delta u(c_4(A,A)) \geq u(c_3(A,B)) + \beta\delta u(c_4(A,B)). \tag{sA2}$$

First, note that (sA1) must bind at an optimum (otherwise, we can raise $c_4(B,B)$, giving the agent a higher utility). Substitute the binding (sA1) in the objective to eliminate $c_4(B,B)$:

$$u(c_1) + \delta u(c_2(A)) + \beta[\delta^2 u(c_3(A,B)) + \delta^3 u(c_4(A,B))] + (\beta - 1)\delta u(c_2(B)).$$

Similarly, (sA2) must bind (otherwise, we can raise $c_4(A,B)$, increasing the agent's utility).

Use the binding (sA2) to rewrite the objective as:

$$u(c_1)+\delta u(c_2(A))+\delta^2 u(c_3(A,A))+\beta\delta^3 u(c_4(A,A))-(1-\beta)[\delta u(c_2(B))+\delta^2 u(c_3(A,B))].$$

Since $\beta < 1$, we should pick $c_2(B)$ and $c_3(AB)$ as small as possible subject to the constraints. Substituting $c_2(B) = c_3(AB) = 0$ back in this expression concludes the proof of Lemma 2. For the proof of Lemma 4, it remains to be verified that the non-lapsing constraints imply perceived non-lapsing constraints if we set $c_2(B) = c_3(A,B) = 0$.

Let $\hat{\mathbf{c}}$ denote a solution to the perceived outside option program, and let $\hat{V}_2^I = u(\hat{c}_2) + \hat{\beta}(\delta u(\hat{c}_3(B)) + \delta^2 u(\hat{c}_4(B)))$. We will use binding ICs constraints to obtain a lower bound on the perceived payoff of keeping the contract and show that is greater than the perceived outside option $\hat{V}_2^I$. We first use the the binding IC for self 2 to rewrite the perceived payoff:

$$\begin{aligned}
&u(c_2(B)) + \hat{\beta}(\delta u(c_3(BB)) + \delta^2 u(c_4(BB))\\
&= u(0) + \frac{\hat{\beta}}{\beta}\beta(\delta u(c_3(B,B)) + \delta^2 u(c_4(B,B))\\
&= u(0) + \frac{\hat{\beta}}{\beta}\left[u(c_2(A)) + \beta(\delta u(c_3(A,B)) + \delta^2 u(c_4(AB))) - u(0)\right]\\
&= \left(1-\frac{\hat{\beta}}{\beta}\right)u(0) + \frac{\hat{\beta}}{\beta}\left[u(c_2(A)) + \beta(\delta u(c_3(A,B)) + \delta^2 u(c_4(AB)))\right],
\end{aligned}$$

where the first equality follows from $c_2(B) = 0$ and the second uses the binding IC constraint (sA1). From the non-lapsing constraint at time 2, we know that $u(c_2(A)) + \beta(\delta u(c_3(A,B)) + \delta^2 u(c_4(AB) \geq V_2^I$, giving a lower bound to the perceived payoff.

$$u(c_2(B)) + \hat{\beta}(\delta u(c_3(BB)) + \delta^2 u(c_4(BB)) \geq \left(1-\frac{\hat{\beta}}{\beta}\right)u(0) + \frac{\hat{\beta}}{\beta}V_2^I.$$

Since V_2^I is the best possible outside option at time 2, in particular, it is greater than or equal to the utility provided by the contract $\hat{c}$, implying

$$
\begin{aligned}
&u(c_2(B)) + \hat{\beta}(\delta u(c_3(BB)) + \delta^2 u(c_4(BB)) \\
&\geq \left(1 - \frac{\hat{\beta}}{\beta}\right) u(0) + \frac{\hat{\beta}}{\beta}\left[u(\hat{c}_2) + \beta(\delta u(\hat{c}_3(B)) + \delta^2 u(\hat{c}_4(B))\right].
\end{aligned}
$$

Rearranging,

$$
\begin{aligned}
&u(c_2(B)) + \hat{\beta}(\delta u(c_3(BB)) + \delta^2 u(c_4(BB)) \\
&= \left(1 - \frac{\hat{\beta}}{\beta}\right) u(0) + \left(\frac{\hat{\beta}}{\beta} - 1\right) u(\hat{c}_2) + \left[u(\hat{c}_2) + \hat{\beta}(\delta u(\hat{c}_3(B)) + \delta^2 u(\hat{c}_4(B))\right] \\
&\geq u(\hat{c}_2) + \hat{\beta}(\delta u(\hat{c}_3(B)) + \delta^2 u(\hat{c}_4(B))) = \hat{V}_2^I,
\end{aligned}
$$

where the inequality comes from $\hat{c}_2 \geq 0$ and $\hat{\beta} \geq \beta$ and the last line comes from the definition of $\hat{V}_2^I$. This shows that the perceived non-lapsing constraints hold.

We next verify that all the perceived choice constraints hold. Notice that

$$
\begin{aligned}
&u(c_3(A,B)) + \hat{\beta}\delta u(c_4(A,B)) = u(0) + \hat{\beta}\delta u(c_4(A,B)) \\
&= \left(1 - \frac{\hat{\beta}}{\beta}\right) u(0) + \frac{\hat{\beta}}{\beta}\left(u(c_3(A,A)) + \beta\delta u(c_4(A,A))\right) \\
&= \left(1 - \frac{\hat{\beta}}{\beta}\right) u(0) + \left(\frac{\hat{\beta}}{\beta} - 1\right) u(c_3(A,A)) + u(c_3(A,A)) + \hat{\beta}\delta u(c_4(A,A)) \\
&\geq u(c_3(A,A)) + \hat{\beta}\delta u(c_4(A,A)), \qquad \text{(sA3)}
\end{aligned}
$$

where the first line uses $u(c_3(A,B)) = 0$, the second line uses the self 3's binding IC constraint (sA2), the third line comes algebraic manipulations, and the last line uses $\hat{\beta} > \beta$

and $c_3(A, A) \geq 0$. Similarly,

$$
\begin{aligned}
& u(c_2(B)) + \hat{\beta}[\delta u(c_3(B, B)) + \delta^2 u(c_4(B, B))] \\
& = u(0) + \hat{\beta}[\delta u(c_3(B, B)) + \delta^2 u(c_4(B, B))] \\
& = (1 - \frac{\hat{\beta}}{\beta}) u(0) + \frac{\hat{\beta}}{\beta} \left[u(c_2(A)) + \beta[\delta u(c_3(A, B)) + \delta^2 u(c_4(A, B))] \right] \\
& \geq u(c_2(A)) + \hat{\beta}[\delta u(c_3(A, B)) + \delta^2 u(c_4(A, B))], \qquad \text{(sA4)}
\end{aligned}
$$

where the first line uses $c_2(B) = 0$, the second line uses the self 2's binding IC constraint (sA1), and the last line uses $\hat{\beta} > \beta$ and $c_2(A) \geq 0$. So the perceived choice constraints hold.

So far, we have shown that $c_2(B) = c_3(AB) = 0$ under the equilibrium contract. We also showed that we can disregard the perceived choice constraints and perceived non-lapsing constraints. Recall that c_t^E denotes the consumption on the equilibrium path at time t. Substituting the binding ICs, the non-lapsing constraints on the equilibrium path can be simplified to $u(c_t^E) + \delta u(c_{t+1}^E) + \cdots + \beta \delta^{t-4} u(c_4^E) \geq V_t^I$.

Therefore, the original program reduces to the auxiliary program:

$$
\max_{(c_1, c_2, c_3, c_4)} u(c_1) + \delta u(c_2) + \delta^2 u(c_3) + \beta \delta^3 u(c_4), \qquad \text{(sA5)}
$$

subject to

$$
\sum_{t=1}^{4} \frac{c_t}{R^{t-1}} = \sum_{t=1}^{4} \frac{w}{R^{t-1}}, \qquad \text{(sA6)}
$$

$$
u(c_t) + \delta u(c_{t+1}) + \cdots + \beta \delta^{4-t} u(c_4) \geq V_t^I, \forall 2 \leq t \leq 4. \qquad \text{(sA7)}
$$

□

Proof of Corollary 1. We can focus on the auxiliary program. Let $x(s_t) \equiv u(c(s_t))$ denote the agent's utility from the consumption he gets in state s_t. We study the dual program:

$$\max_{\{x(s_t)\}} \sum_{t=1}^{T} \sum_{s_t \in \mathbb{S}_t} p(s_t|s_1) \frac{w(s_t) - u^{-1}(x(s_t))}{R^{t-1}}, \tag{sA8}$$

subject to

$$\sum_{t=1}^{T-1} \sum_{s_t \in \mathbb{S}_t} \delta^{t-1} p(s_t|s_1) x(s_t) + \beta \sum_{s_T \in \mathbb{S}_T} \delta^{T-1} p(s_T|s_1) x(s_T) \geq \underline{u}. \tag{sA9}$$

This program corresponds to the maximization of a strictly concave function over a convex set, so that, by the Theorem of the Maximum, the solution is unique. Moreover, the consumption path is continuous in $\beta \in (0,1]$. Finally, the program does not involve $\hat{\beta}$, so the consumption path is not a function of the consumer's naiveté.

Once we pin down the unique consumption path, the baseline options are either zero or determined by the binding IC constraints, which do not depend on $\hat{\beta}$ (see the proof of Lemma 2). So the equilibrium consumption vector is not a function of the consumer's naiveté.

□

Proof of Claim 1 (from the proof of Theorem 3). First, the time consistent agent's welfare is exactly given by the outside option,

$$\hat{W}_T^C(\underline{\mathbf{c}}) = E \sum_{t=1}^{T} \delta^{t-1} u(\underline{c}(s_t)).$$

The limit of $\hat{W}_T^C(\underline{\mathbf{c}})$ exists by the root test:

$$\limsup_{T \nearrow \infty} \sqrt[T]{\delta^{T-1} |u(\underline{c}(s_T))|} \leq \delta < 1.$$

Second, we show the limit of $\Pi_T^C(\underline{\mathbf{c}})$ exists using the Cauchy convergence criterion. Specifically, we claim that for sufficiently large T,

$$E\frac{w(s_T) - \underline{c}(s_T)}{R^{T-1}} \leq \Pi_T^C(\underline{\mathbf{c}}) - \Pi_{T-1}^C(\underline{\mathbf{c}}) \leq E\frac{w(s_T)}{R^{T-1}} + \frac{2\delta^{T-1}\max\{|u(\cdot)|\}}{u'(K)}. \tag{sA10}$$

The claim follows from a revealed-preference argument. Suppose $(c_1', \cdots, c'(s_{T-1}))$ solves the program $\Pi_{T-1}^C(\underline{\mathbf{c}})$. Then $(c_1', \cdots, c'(s_{T-1}), \underline{c}(s_T))$ is in the feasible set of the program $\Pi_T^C(\underline{\mathbf{c}})$. By the revealed-preference argument, it immediately follows that

$$\Pi_T^C(\underline{\mathbf{c}}) \geq E\frac{w(s_T) - \underline{c}(s_T)}{R^{T-1}} + \Pi_{T-1}^C(\underline{\mathbf{c}}).$$

Suppose $(c_1^*, \cdots, c^*(s_T))$ solves the program $\Pi_T^C(\underline{\mathbf{c}})$. We show that

$$\left(c_1^* + \frac{2\delta^{T-1}\max\{|u(\cdot)|\}}{u'(K)}, c^*(s_2), \cdots, c^*(s_{T-1})\right) \tag{sA11}$$

is in the feasible set of the program $\Pi_{T-1}^C(\underline{\mathbf{c}})$. To see that, note that from the Lagrange's Mean Value Theorem, it follows that

$$u\left(c_1^* + \frac{2\delta^{T-1}\max\{|u(\cdot)|\}}{u'(K)}\right) - u(c_1^*) = u'(\xi)\frac{2\delta^{T-1}\max\{|u(\cdot)|\}}{u'(K)}, \tag{sA12}$$

where $\xi \in (c_1^*, c_1^* + \frac{2\delta^{T-1}\max\{|u(\cdot)|\}}{u'(K)})$. For sufficiently large T, $c_1^* + \frac{2\delta^{T-1}\max\{|u(\cdot)|\}}{u'(K)} < K$. So $u'(\xi) \geq u'(K)$. Going back to equation (sA12) leads to

$$u\left(c_1^* + \frac{2\delta^{T-1}\max\{|u(\cdot)|\}}{u'(K)}\right) - u(c_1^*) \geq 2\delta^{T-1}\max\{|u(\cdot)|\}. \tag{sA13}$$

Then,

$$
\begin{aligned}
& u\left(c_1^* + \frac{2\delta^{T-1}\max\{|u(\cdot)|\}}{u'(K)}\right) + E\sum_{t=2}^{T-1}\delta^{t-1}u(c^*(s_t)) \\
& \geq 2\delta^{T-1}\max\{|u(\cdot)|\} + E\sum_{t=1}^{T-1}\delta^{t-1}u(c^*(s_t)) \\
& \geq 2\delta^{T-1}\max\{|u(\cdot)|\} + E\sum_{t=1}^{T}\delta^{t-1}u(\underline{c}_t) - E\delta^{T-1}u(c^*(s_T)) \\
& = E\sum_{t=1}^{T-1}\delta^{t-1}u(\underline{c}_t) + \left(\delta^{T-1}\max\{|u(\cdot)|\} - E\delta^{T-1}u(c^*(s_T))\right) + \left(\delta^{T-1}\max\{|u(\cdot)|\} + E\delta^{T-1}u(\underline{c}_T)\right) \\
& \geq E\sum_{t=1}^{T-1}\delta^{t-1}u(\underline{c}_t),
\end{aligned}
$$

where the first inequality comes from (sA13), the second comes from noting that $(c_1^*, \cdots, c^*(s_T))$ solves program $\Pi_T^C(\underline{\mathbf{c}})$, the equality comes from algebraic manipulations, and the last step uses the boundedness of u. So we have shown that (sA11) is in the feasible set of $\Pi_{T-1}^C(\underline{c})$. A revealed-preference argument implies that

$$
\Pi_{T-1}^C(\underline{\mathbf{c}}) \geq E\sum_{t=1}^{T-1}\frac{w(s_t) - c^*(s_t)}{R^{t-1}} - \frac{2\delta^{T-1}\max\{|u(\cdot)|\}}{u'(K)}.
$$

Recall that $\Pi_T^C(\underline{\mathbf{c}}) = E\sum_{t=1}^{T}\frac{w(s_t)-c^*(s_t)}{R^{t-1}}$. Substituting it back to the previous inequality, we obtain

$$
\Pi_{T-1}^C(\underline{\mathbf{c}}) \geq \Pi_T^C(\underline{\mathbf{c}}) - E\frac{w(s_T) - c^*(s_T)}{R^{T-1}} - \frac{2\delta^{T-1}\max\{|u(\cdot)|\}}{u'(K)},
$$

establishing the right-hand-side of (sA10) because of $c^*(s_T) \geq 0$. Since $E\sum_{t=1}^{T}\frac{w(s_T)-\underline{c}(s_T)}{R^{T-1}}$ exists and $\delta < 1$, for $\forall\epsilon$, we can find T_0 such that $\forall T_1, T_2 > T_0$, $|\Pi_{T_1}^C(\underline{\mathbf{c}}) - \Pi_{T_2}^C(\underline{\mathbf{c}})| < \epsilon$. This establishes that $\{\Pi_T^C(\underline{\mathbf{c}})\}$ satisfies the Cauchy convergence criterion, therefore the limit exists. □

Proof of Proposition 2. Since c^* maximizes the welfare function $W_T^H(c)$, it immediately follows that $W_T^{H,I} \leq W_T^*, \forall T$. Thus,

$$\limsup_{T \nearrow \infty} \frac{W_T^{H,I} - W_T^*}{T} \leq 0. \tag{sA14}$$

Denote $d_{t,T} = \frac{D_{T-1}}{D_{T-t}}, \forall t = 1, \cdots, T$. The objective function in the naive agent's auxiliary program becomes

$$\sum_{t=1}^{T} d_{t,T} u(c_t). \tag{sA15}$$

It follows that

$$\begin{aligned} W_T^{H,I} = \sum_{t=1}^{T} u(c_t^H) &= \sum_{t=1}^{T} [d_{t,T} u(c_t^H) + (1 - d_{t,T}) u(c_t^H)] \\ &\geq \sum_{t=1}^{T} [d_{t,T} u(c_t^*) + (1 - d_{t,T}) u(c_t^H)], \end{aligned}$$

where the first line comes from the definition and algebraic manipulations and the last step comes from the fact that c^H maximizes (sA15) and that c^* is feasible. Rearranging,

$$\begin{aligned} W_T^{H,I} &\geq \sum_{t=1}^{T} \left[u(c_t^*) + (1 - d_{t,T}) [u(c_t^H) - u(c_t^*)] \right] \\ &= \sum_{t=1}^{T} \left[u(c_t^*) + \frac{k(t-1)}{1 + k(T-1)} [u(c_t^H) - u(c_t^*)] \right] \\ &= W_T^* + \sum_{t=1}^{T} \frac{k(t-1)}{1 + k(T-1)} [u(c_t^H) - u(c_t^*)], \end{aligned} \tag{sA16}$$

where the first line comes from algebraic manipulations, the second line uses the definition of $d_{t,T}$, and the last line comes from the definition of W_T^*.

We next show a series of lemmas to bound the second term. Let λ^H denote the Lagrangian multiplier from the zero-profit condition in the naive agent's program, and let λ^* denote the Lagrangian multiplier from the time-consistent agent's program. Note that the solution must be interior solution since $\lim_{c \searrow 0} u'(0) = +\infty$.

Lemma 10. *There exist* $\underline{\lambda}, \overline{\lambda} \in (0, +\infty)$ *such that*

$$\underline{\lambda} \leq \min(\lambda^H, \lambda^*) \leq \max(\lambda^H, \lambda^*) \leq \overline{\lambda}.$$

Proof. From the first-order-condition, we know that

$$\lambda^H = u'(c_1^H), \lambda^* = u'(c_1^*).$$

Note that the first period consumption must be between 0 and $\sum_{t=1}^{\infty} \frac{w}{R^{t-1}} = \frac{w}{1-R}$. The lemma follows immediately by letting $\overline{\lambda} = u'(0)$ and $\underline{\lambda} = u'\left(\frac{w}{1-R}\right)$. □

Lemma 11. *There exists a constant* $A > 0$ *such that* $|t(u(c_t^H) - u(c_t^*))| < A, \forall t, \forall T.$

Proof. From the first-order-condition, we know that

$$\frac{\lambda^H d_{t,T}}{R^{t-1}} = u'(c_t^H), \frac{\lambda^*}{R^{t-1}} = u'(c_t^*)$$

Denote $g(\cdot) = (u')^{-1}(\cdot)$. Inverting above equations to solve for c_t^H and c_t^*,

$$c_t^H = g\left(\frac{\lambda^H d_{t,T}}{R^{t-1}}\right), c_t^* = g\left(\frac{\lambda^*}{R^{t-1}}\right).$$

Note that $\frac{du(g(x))}{dx} = \frac{x}{u''(g(x))}$. Applying Lagrangian Mean Value Theorem, there exists η, where $\frac{\min(\lambda^*, \lambda^H d_{t,T})}{R^{t-1}} \leq \eta \leq \frac{\max(\lambda^*, \lambda^H d_{t,T})}{R^{t-1}}$, such that

$$|t(u(c_t^H) - u(c_t^*))| = t\left|u\left(g\left(\frac{\lambda^H d_{t,T}}{R^{t-1}}\right)\right) - u\left(g\left(\frac{\lambda^*}{R^{t-1}}\right)\right)\right| \tag{sA17}$$

$$= t\left|\frac{\eta}{u''(g(\eta))}\left(\frac{\lambda^H d_{t,T}}{R^{t-1}} - \frac{\lambda^*}{R^{t-1}}\right)\right|. \tag{sA18}$$

Using a change of variable $x = \frac{1}{R^{t-1}}$, then

$$x\underline{\lambda} d_{t,T} \leq x \min(\lambda^*, \lambda^H d_{t,T}) \leq \eta \leq x \max(\lambda^*, \lambda^H d_{t,T}) \leq x\overline{\lambda}.$$

So $x \geq \frac{\eta}{\overline{\lambda}}$. Note that $d_{t,T} = \frac{1+k(T-t)}{1+k(T-1)} \geq \frac{1}{1+k(t-1)}$. So,

$$x \leq \frac{\eta}{\underline{\lambda} d_{t,T}} \leq \frac{\eta(1+k(t-1))}{\underline{\lambda}} = \frac{\eta(1-k\frac{\log(x)}{\log R})}{\underline{\lambda}} \leq \frac{\eta(1-k\frac{\log(\eta)-\log(\overline{\lambda})}{\log R})}{\underline{\lambda}}. \tag{sA19}$$

We can rewrite (sA18) as

$$\begin{aligned}
|t(u(c_t^H) - u(c_t^*))| &\leq \left(-\frac{\log x}{\log R} + 1\right) \left|\frac{\eta}{u''(g(\eta))}\right| \frac{2\overline{\lambda}}{R^{t-1}} \\
&= \left(-\frac{\log x}{\log R} + 1\right) \left|\frac{\eta}{u''(g(\eta))}\right| 2\overline{\lambda} x \\
&\leq 2\frac{\overline{\lambda}}{\underline{\lambda}} \left(-\frac{\log \eta - \log(\overline{\lambda})}{\log R} + 1\right) \left(1 - k\frac{\log \eta - \log \overline{\lambda}}{\log R}\right) \frac{\eta^2}{|u''(g(\eta))|} \\
&\leq constant * \frac{(\log \eta)^2 \eta^2}{|u''(g(\eta))|},
\end{aligned}$$

where the first line uses $t = -\frac{\log x}{\log R} + 1$ and Lemma 10, the second line uses $x = \frac{1}{R^{t-1}}$, the third uses (sA19), and the last line collects the first-order terms. Let $\xi = g(\eta)$ and use Assumption 1, so there exists $A > 0$ such that $|t(u(c_t^H) - u(c_t^*))| < A$. □

Lemma 12. $\sum_{t=1}^{T} \frac{1}{t} \geq \log(T)$ *for any* $T \geq 1$.

Proof. Note that $\log(t+1) - \log(t) = \int_t^{t+1} \frac{1}{\theta} d\theta \leq \frac{1}{t}$. Sum over t from 1 to $(T-1)$ to obtain: $\log(T) \leq \sum_{t=1}^{T-1} \frac{1}{t} \leq \sum_{t=1}^{T} \frac{1}{t}$. □

Lemma 13. *There exists a constant* $A' > 0$ *such that*

$$\sum_{t=1}^{T} \frac{k(t-1)}{1+k(T-1)} |u(c_t^H) - u(c_t^*)| < A', \forall T$$

Proof. Using Lemma 11, it follows that

$$\begin{aligned}\sum_{t=1}^{T}\frac{k(t-1)}{1+k(T-1)}[u(c_t^H)-u(c_t^*)] &\leq \sum_{t=1}^{T}\frac{k(t-1)}{1+k(T-1)}\frac{A}{t}\\ &\leq \sum_{t=1}^{T}\frac{k}{1+k(T-1)}A+\sum_{t=1}^{T}\frac{-k}{1+k(T-1)}A\frac{1}{t}\\ &\leq \frac{kAT}{1+k(T-1)}+\frac{-k}{1+k(T-1)}A\log(T),\end{aligned}$$

where the first line comes from the lemma 11, the second line comes from algebraic manipulations, and the last line comes from $k \geq 0$ and lemma 12. Note that as $T \nearrow \infty$, the first term converges to A, and the second term converges to 0. So there exists a constant $A' > 0$ such that

$$\sum_{t=1}^{T}\frac{k(t-1)}{1+k(T-1)}[u(c_t^H)-u(c_t^*)] < A', \forall T.$$

□

Returning to (sA16), we have

$$\liminf_{T\nearrow\infty}\frac{W_T^{H,I}-W_T^*}{T} \geq -\liminf_{T\nearrow\infty}\frac{A'}{T}=0. \tag{sA20}$$

Together with (sA14), it implies that $\lim_{T\nearrow\infty}\frac{W_T^{H,I}-W_T^*}{T}$ exists, and

$$\lim_{T\nearrow\infty}\frac{W_T^{H,I}-W_T^*}{T}=0.$$

□

Proof of Proposition 4. It is easy to construct off-path beliefs that support the full-information allocation as an equilibrium. We need to show that no other allocation can be supported as an equilibrium. Suppose there exists a type $\hat{\beta}_0$ that does not pick the full information contract in equilibrium. There are two possibilities: (i) $\hat{\beta}_0$ is separated in equilibrium (i.e., no other type picks the same contract at $\hat{\beta}_0$), or (ii) $\hat{\beta}_0$ is pooled in equilibrium (i.e. there exists another type that picks the same contract as $\hat{\beta}_0$).

Consider case (i) first. Since $\hat{\beta}_0$ is the only type picking its contract, that contract must satisfy IC, PC, and zero profits. Recall that the full information contract is the unique contract that maximizes self 1's perceived utility subject to IC, PC, and zero profits. Consider a deviation in which type $\hat{\beta}_0$ offers the full-information contract in all histories except in period 1, where it offers a slightly lower consumption than with full information. Note that lowering c_1 does not affect IC and PC and, by taking c_1 arbitrarily close to the full-information consumption, we ensure that the consumer gets a strictly higher perceived utility while leaving strictly positive profits to the firm, contradicting the assumption that the original allocation was part of an equilibrium.

Next consider case (ii), so there are at least two types pooled at a contract different from the full information contract. If the firm breaks even on each consumer, then by the same argument as before, all consumers would strictly benefit from deviating to offering the full information contract (with a slightly lower c_1), which also gives strictly positive profits for the firm. If instead there is a cross subsidy between types, a type that is providing a positive profit can strictly benefit from deviating to the full information contract (with a slightly lower c_1). Moreover, by taking c_1 close enough to one in the full-information contract (which maximizes the perceived utility and leaves zero profits), we ensure that deviation is profitable. □

Proof of Proposition 5. Suppose we have an equilibrium in which at least one naive type does not pick the full-information contract. Using the same argument as in Proposition 4, that type cannot be separated or pooled with other naive types only. Therefore, the only remaining case is one where at least one naive type pools with the sophisticated type.

But note that the contract that a sophisticated type would offer under full information offers a fixed consumption in each period (no alternative options), maximizing his perceived utility at time 1 under the zero profits constraint. Therefore, he must be cross subsidized in order to choose another contract (i.e., the firm must make strictly negative profits from serving him). But since the firm would not accept a contract that makes negative profits, this means that the firm must make strictly positive profits on some naive type that is pooling with the sophisticated type. But then this naive type would strictly profit from deviating to full information contract, which maximizes his perceived utility at time 1 subject to the zero profits constraint. □

Proof of Lemma 5. Without loss of generality, we use the following normalization $u(0) = 0$ in our analysis below. The proof follows by contradiction. Suppose there is an equilibrium in which two types, β_L and $\beta_H > \beta_L$, offer their full information contracts, $\mathcal{C}^L$ and $\mathcal{C}^H$. We show that these contracts cannot be part of an equilibrium when T is large enough since type β_H would deviate and pick $\mathcal{C}^L$, leaving the firm with negative profits.

Note that, by the binding IC constraint for type β_L, if type β_H picks $\mathcal{C}^L$, he ends up choosing B rather than A. In this case, the firm offering $\mathcal{C}^L$ makes negative profits. To see this, suppose instead that the firm makes a non-negative profit from this contract. But this would mean that the non-flexible contract that gives only the baseline consumption would also solve type β_L's program, which contradicts Corollary 1.

Type β_H's perceived utility from $\mathcal{C}^L$ equals:

$$
\begin{aligned}
& u(c_1^L) + \beta_H \sum_{t=2}^{T} \delta^{t-1} u(c_t^L(B, \cdots, B)) \\
&= \frac{\beta_L - \beta_H}{\beta_L} u(c_1^L) + \frac{\beta_H}{\beta_L} u(c_1^L) + \beta_H \sum_{t=2}^{T} \delta^{t-1} u(c_t^L(B, \cdots, B)) \\
&= \frac{\beta_L - \beta_H}{\beta_L} u(c_1^L) + \frac{\beta_H}{\beta_L} \left[u(c_1^L) + \beta_L \sum_{t=2}^{T} \delta^{t-1} u(c_t^L(B, \cdots, B)) \right] \\
&= \frac{\beta_L - \beta_H}{\beta_L} u(c_1^L) + \frac{\beta_H}{\beta_L} \left[\sum_{t=1}^{T-1} \delta^{t-1} u(c_t^L(A, \cdots, A)) + \beta_L \delta^{T-1} u(c_T^L(A, \cdots, A)) \right] \\
&= \frac{\beta_L - \beta_H}{\beta_L} u(c_1^L) + \frac{\beta_H}{\beta_L} \sum_{t=1}^{T-1} \delta^{t-1} u(c_t^L(A, \cdots, A)) + \beta_H \delta^{T-1} u(c_T^L(A, \cdots, A)) \\
&= u(c_1^L) + \frac{\beta_H}{\beta_L} \sum_{t=2}^{T-1} \delta^{t-1} u(c_t^L(A, \cdots, A)) + \beta_H \delta^{T-1} u(c_T^L(A, \cdots, A))
\end{aligned}
$$

$$
\begin{aligned}
&> \sum_{t=1}^{T-1} \delta^{t-1} u(c_t^H(A, \cdots, A)) + \beta_H \delta^{T-1} u(c_T^H(A, \cdots, A)) \\
&= u(c_1^H) + \beta_H \left[\sum_{t=2}^{T} \delta^{t-1} u(c_t^H(B, \cdots, B)) \right].
\end{aligned}
$$

where the second, third, fifth, and sixth lines follow from algebraic manipulations, the fourth line substitutes the binding (IC) for the low type, and the last line uses the binding (IC) for the high type. The strict inequality on the seventh line uses the following facts: $\beta_H > \beta_L$, $u_t(A, A..., A) \geq 0$ with strict inequality for at least one t, and, from Theorem 1, the welfare of time-inconsistent consumers converge to the welfare of time-consistent consumers

$$
\lim_{T \nearrow \infty} \sum_{t=1}^{T-1} \delta^{t-1} u(c_t^L) + \beta_H \delta^{T-1} u(c_T^L) = \lim_{T \nearrow \infty} \sum_{t=1}^{T-1} \delta^{t-1} u(c_t^H) + \beta_H \delta^{T-1} u(c_T^H).
$$

Therefore, for T sufficiently large, the β_H consumer would have an incentive to deviate and choose β_L consumer's full-information contract while taking the baseline option. $\square$

Proof of Lemma 6. We argue by contradiction. Fix an equilibrium in which β_H does not get his full-information contract. First, suppose that firms make non-negative profits from β_H. Suppose β_H deviates and offers his full-information contract. By a single-crossing argument, if type β_L got β_H's full-information contract, he would always choose option A, so the firm would break even on both types under type β_H's full-information contract. Since the full-information contract maximizes β_H's perceived utility among those that make zero profits, β_H has an incentive to deviate to it.

Suppose, instead, that the firm makes strictly negative profits on type β_H. Then firm optimality requires that both types pool on the same contract $\mathcal{C}$ and the firm makes strictly positive profits on type β_L. To generate different profits, these two types must be getting different allocations on the equilibrium path.

We will construct a deviation contract $\mathcal{C}(\epsilon)$ such that whenever the β_H consumer weakly benefits from the deviation, the β_L consumer strictly benefits from the deviation. By D1 criteria, we should assign zero weight to the type β_H and all the weight to the type β_L consumer. Given that firms make positive profits from the β_L consumer's equilibrium contract, firms would charge a price such that the β_L consumer are better off with $\mathcal{C}(\epsilon)$ than the contract $\mathcal{C}$, a contradiction. Since both types have the same naiveté parameter $\hat{\beta}$, they both believe they will choose the same options. Let $(\hat{c}_2, \cdots, \hat{c}_T)$ denote their perceived consumption stream. Construct a perturbation of the equilibrium contract, $\mathcal{C}(\epsilon)$, by decreasing the last-period perceived consumption by ϵ and adjusting the other options so that (IC) and (PC) hold for both types. Upon observing contract $\mathcal{C}(\epsilon)$, the firm must assign full weight to type β_L. This is because whenever β_H benefits from deviating to this contract (i.e., when the firm's price is lower than $-\beta_H \delta^{T-1} \frac{u'(c_T)}{u'(c_1)} \epsilon$, β_L also benefits from this deviation (i.e., when the firm's price is lower than $-\beta_L \delta^{T-1} \frac{u'(c_T)}{u'(c_1)} \epsilon$. Therefore, this candidate equilibrium does not satisfy D1. □

Proof of Lemma 7. To show that the proposed equilibrium survives D1, we show that if β_L can benefit from a deviation to $\mathcal{C}'$, then β_H strictly benefits from the deviation as well.

Recall that they have the same perceived time-consistency parameter $\hat{\beta}$, so their perceived consumption from the contract $\mathcal{C}'$ are the same, denoted as $(c'_1, c'_2, \cdots, c'_T)$. Suppose β_L can benefit from the deviation:

$$u(\bar{c}_1) + \beta_L \sum_{t=2}^{T} \delta^{t-1} u(\bar{c}_t(B, \cdots, B)) < u(c'_1) + \beta_L \sum_{t=2}^{T} \delta^{t-1} u(c'_t). \tag{sA21}$$

By a single-crossing argument, since $\bar{c}$ solves β_L's program (and therefore his IC must bind), type β_H's IC cannot hold:

$$u(c_1^H) + \beta_H \sum_{t=2}^{T} \delta^{t-1} u(c_t^H(B, \cdots, B)) < u(c'_1) + \beta_H \sum_{t=2}^{T} \delta^{t-1} u(c'_t), \tag{sA22}$$

so β_H also benefits from this deviation. According to D1, we must assign zero weight on β_L and full weight on β_H. Because β_H gets his full-information contract in any equilibrium satisfying D1 (Lemma 6), we must have

$$u(c_1^H) + \beta_H \sum_{t=2}^{T} \delta^{t-1} u(c_t^H(B, \cdots, B)) \geq u(c_1') + \beta_H \sum_{t=2}^{T} \delta^{t-1} u(c_t'), \tag{sA23}$$

a contradiction to (sA22). So we have

$$u(\bar{c}_1) + \beta_L \sum_{t=2}^{T} \delta^{t-1} u(\bar{c}_t(B, \cdots, B)) \geq u(c_1') + \beta_L \sum_{t=2}^{T} \delta^{t-1} u(c_t'), \tag{sA24}$$

showing that β_L does not have a profitable deviation and the proposed equilibrium survives D1.

Next, we show that in any equilibrium satisfying D1, the consumption path corresponds to the least costly separating allocation. Suppose there exists another equilibrium that survives D1. As we showed above, β_H gets his full-information contract. Let $\mathcal{C}'$ denote β_L's equilibrium contract, and the contract is different from the leastly costly separation allocation (A1). Suppose β_L deviates and offers a contract that coincides with the solution to (A1) except that it reduces consumption in the first period by a small $\epsilon > 0$. By the IC constraint, the β_H consumer is strictly worse off by choosing this new contract instead of his full-information contract. By D1, firms must assign full weight to β_L. By choosing ϵ small enough, β_L strictly benefits from the deviation. $\square$

Proof of Proposition 6. From the previous lemmas, the equilibrium is given by the least-costly separation. The equilibrium-path consumption for the low type solves the following program:

$$\max u(c_1) + l(c_1),$$

subject to

$$u(c_1) + \frac{\beta_H}{\beta_L} l(c_1) = V^H, \tag{sA25}$$

where $l(\cdot)$ is defined as

$$l(c_1) = \max_{(c_2, c_3, \cdots, c_T)} \sum_{t=2}^{T-1} \delta^{t-1} u(c_t) + \beta \delta^{T-1} u(c_T).$$

subject to

$$\sum_{t=2}^{T} \frac{c_t}{R^{t-1}} = \sum_{t=1}^{T} \frac{w}{R^{t-1}} - c_1.$$

Note that by substituting (sA25) to the objective function, maximizing $u(c_1) + l(c_1)$ is equivalent to maximizing c_1. If β_L's full-information contract cannot be sustained in a equilibrium (as must be the case if T is large), it means that

$$u(c_1^L) + \frac{\beta_H}{\beta_L} l(c_1^L) > V^H. \tag{sA26}$$

Evaluating c_1 at $\sum_{t=1}^{T} \frac{w}{R^{t-1}}$ implies that the

$$u\left(\sum_{t=1}^{T} \frac{w}{R^{t-1}}\right) + \frac{\beta_H}{\beta_L} \delta l \left(\sum_{t=1}^{T} \frac{w}{R^{t-1}}\right) = u\left(\sum_{t=1}^{T} \frac{w}{R^{t-1}}\right) < V^H. \tag{sA27}$$

By the intermediate value theorem, it follows that the maximal root of (sA25) must be greater than the first period consumption in the full-information contract: $\bar{c}_1 > c_1^L$. This completes the first part of the proposition.

We next show that the welfare loss must be bounded below away from 0. We argue by contradiction. Suppose there exists a subsequence $\{T_n : n \in \mathbb{N}\}$ such that $\lim_{n \nearrow \infty}(W_{T_n}^C - W_{T_n}^L) = 0$. To be clear that our variables now depends on T_n, we write variables as a function of T_n. Note that

$$\begin{aligned}
\lim_{n\nearrow\infty}(W^C_{T_n} - W^L_{T_n}) &= \lim_{n\nearrow\infty}(V^H(T_n) - W^L_{T_n}) \\
&= \lim_{n\nearrow\infty}\left(V^H(T_n) - u(\overline{c}_1(T_n)) - l(\overline{c}_1(T_n))\right) \\
&= \lim_{n\nearrow\infty}\left(V^H(T_n) - u(\overline{c}_1(T_n)) - (V^H(T_n) - u(\overline{c}_1(T_n)))\frac{\beta_L}{\beta_H}\right) \\
&= \lim_{n\nearrow\infty}\left(V^H(T_n) - u(\overline{c}_1(T_n))\right)\left(1 - \frac{\beta_L}{\beta_H}\right),
\end{aligned}$$

where the first equality comes from the vanishing inefficiency result for the β_H consumer, the second equality comes from the definition of $l(c_1)$ and $(1-\beta)\delta^{T-1}u(\overline{c}(s_T)) \to 0$, the third equality comes from (sA25), and the fourth equality comes from algebraic manipulations.

It follows that $\lim_{n\nearrow\infty}\left(V^H(T_n) - u(\overline{c}_1(T_n))\right) = 0$. From the vanishing inefficiency result for the β_H consumer, it implies that

$$\lim_{n\nearrow\infty}\left(W^L_{T_n} - u(\overline{c}_1(T_n))\right) = \lim_{n\nearrow\infty}\left(W^C_{T_n} - u(\overline{c}_1(T_n))\right) = \lim_{n\nearrow\infty}\left(V^H(T_n) - u(\overline{c}_1(T_n))\right) = 0. \tag{sA28}$$

Note that $W^L_{T_n} = \sum_{t=1}^{T_n}\delta^{t-1}u(\overline{c}_t(T_n))$. We obtain

$$\lim_{n\nearrow\infty}\sum_{t=2}^{T_n}\delta^{t-1}u(\overline{c}_t(T_n)) = 0.$$

Recall that we normalize $u(0) = 0$, so $u(c) \geq 0, \forall c \geq 0$. We must have $\lim_{n\nearrow\infty}\overline{c}_t(T_n) = 0, \forall t$. By the zero-profits condition, the β_L consumer consumes everything in the first period in the limit: $\lim_{n\nearrow\infty}\overline{c}_1(T_n) = \sum_{t=1}^{\infty}\frac{w}{R^{t-1}}$. This consumption stream cannot achieve the first-best welfare (i.e., W^C_T), as shifting some consumption to future periods can strictly improves welfare since $\lim_{c\searrow 0}u'(c) = +\infty$. Specifically, fix a small $\epsilon_0 > 0$, it is straightforward to show that

$$u\left(\sum_{t=1}^{\infty}\frac{w}{R^{t-1}}\right) < u\left(\sum_{t=1}^{\infty}\frac{w}{R^{t-1}} - \epsilon_0\right) + \delta u(R\epsilon_0) \leq \lim_{T\nearrow\infty} W_T^C.$$

This is a contradiction to (sA28) that $\lim_{n\nearrow\infty}\left(W_{T_n}^C - u(\overline{c}_1(T_n))\right) = 0$.

So the welfare loss does not vanish as the contracting horizon grows: $\liminf_{T\nearrow\infty}(W_T^C - W_T^L) > 0$. □

Proof of Lemma 8. To prove the lemma, we argue by contradiction. There exist two option history paths of consumption stream starting with (s_t, h^t) that have different expected present discounted values. Without loss of generality, assume that one path, denoted as $\hat{\mathbf{c}}$, has a higher expected present value than the other path, denoted as $\tilde{\mathbf{c}}$.

We note that since $\mathbf{c}$ is the equilibrium consumption vector, it must satisfy the *no additional contracting constraints*. Given that the present value of $\hat{\mathbf{c}}$ is higher than the present value of $\tilde{\mathbf{c}}$. There are two possibilities, either $\tilde{\mathbf{c}}$ starts with the baseline option or $\tilde{\mathbf{c}}$ starts with the alternative option. In either case, we show that the no additional contracting constraints would be violated. First, suppose $\tilde{\mathbf{c}}$ starts with the baseline option. In this case, the baseline option would not be the consumer's perceived consumption, because the consumer perceives that he has an incentive to recontract with another firm, who can give the consumer slightly higher consumption in the baseline option. Specifically, consider another contract $\mathbf{c}'$, which has the same term as $\mathbf{c}$ except that we increase ϵ in the consumption in the baseline option of $\tilde{\mathbf{c}}$. Similarly, if $\tilde{\mathbf{c}}$ starts with the alternative option, the consumer can recontract with another firm, who gives him slightly higher consumption in the alternative option. □

Proof of Proposition 7. Note first that uncertainty over states plays no role in the program with non-exclusive contracts. Starting from any allocation in which consumption within a period is random, the agent increase his perceived utility by signing a contract with another firm to smooth consumption in that period. So we can without loss of generality substitute each period's income by its expected value. Using Lemma 8, we find that the program with non-exclusive contracts becomes identical to the consumption-savings problem. □

Proof of Proposition 8. Consider a problem with a sophisticated consumer who has the commitment power and whose time-consistency parameter is $\frac{\beta}{\hat{\beta}}$. Without loss of generality, we assume that there is no uncertainty. Recall that a_1 is the PDV of income. The sophisticate's program is

$$c^S = \max_{\{c(\cdot)\}} u(c_1) + \frac{\beta}{\hat{\beta}} \sum_{t=2}^{T} \delta^{t-1} u(c_t),$$

subject to

$$\sum_{t=1}^{T} \frac{c_t}{R^{t-1}} = a_1.$$

We claim that the welfare in the above program is an upper bound of the welfare in the consumption-savings problem for the naive consumer. Let $\bar{c}_1$ denote the first-period consumption in the consumption-savings problem. We will show that the naive agent consumes strictly more than the sophisticated agent: $\bar{c}_1 \geq c_1^S$.

The proof proceeds through four lemmas. The first one adapts arguments from Harris and Laibson (2001).

Lemma 14. *The perceived consumption functions* $(\hat{c}_2(\cdot), \cdots, \hat{c}_T(\cdot))$ *satisfy:*

$$(\delta R)^{t-1} u'(\hat{c}_t) = (\delta R)^t u'(\hat{c}_{t+1}) \left[1 - \hat{c}'_{t+1}(a_{t+1}) + \hat{\beta} \hat{c}'_{t+1}(a_{t+1}))\right], \forall 1 < t < T,$$

where $a_{t+1} = R(a_t - \hat{c}_t(a_t))$.

Proof. The proof follows by induction, starting at period $T-1$. The last period consumption is $c_T(a_T) = a_T$. Consumption in the penultimate period is:

$$\hat{c}_{T-1}(a_{T-1}) = \arg\max_{\tilde{c}} \{u(\tilde{c}) + \hat{\beta}\delta u(\hat{c}_T(a_T)) \text{ subject to } a_T = R(a_{T-1} - \tilde{c})\}.$$

Since $\lim_{c \searrow 0} u'(c) = +\infty$, the unique solution must be interior and satisfy the FOC:

$$u'(\hat{c}_{T-1}) = \hat{\beta}\delta R u'(\hat{c}_T).$$

Since $\hat{c}'_T(a_T) = 1$, the statement in the lemma holds for $t = T-1$.

Moving to the induction step, suppose the statement holds for $\tau < T$ and recall that:

$$\hat{c}_\tau(a_\tau) = \arg\max_{\tilde{c}} \{u(\tilde{c}) + \hat{\beta} \sum_{t=\tau+1}^{T} \delta^{t-\tau} u(\hat{c}_t(a_t)) \text{ subject to (B2), (B3), and (B4)}\}.$$

The unique solution must be interior and satisfy the FOC:

$$u'(\hat{c}_\tau) + \hat{\beta} \sum_{t=\tau+1}^{T} \delta^{t-\tau} u'(\hat{c}_t(a_t)) \frac{\partial \hat{c}_t(a_t)}{\partial \hat{c}_\tau} = 0.$$

Substitute

$$\begin{aligned} \frac{\partial \hat{c}_t(a_t)}{\partial \hat{c}_\tau} &= \hat{c}'_t(a_t) \frac{\partial a_t}{\partial \hat{c}_\tau} = \hat{c}'_t(a_t) \frac{\partial a_t}{\partial a_{t-1}} \cdots \frac{\partial a_{\tau+1}}{\partial \hat{c}_\tau} \\ &= -R^{t-\tau} \hat{c}'_t(a_t)(1 - \hat{c}'_{t-1}(a_{t-1})) \cdots (1 - \hat{c}'_{\tau+1}(a_{\tau+1})), \end{aligned}$$

to rewrite the FOC as:

$$u'(\hat{c}_\tau) = \hat{\beta} \sum_{t=\tau+1}^{T} (\delta R)^{t-\tau} u'(\hat{c}_t(a_t)) \hat{c}'_t(a_t)(1 - \hat{c}'_{t-1}(a_{t-1})) \cdots (1 - \hat{c}'_{\tau+1}(a_{\tau+1})).$$

The FOC at $\tau + 1$ is:

$$u'(\hat{c}_{\tau+1}) = \hat{\beta} \sum_{t=\tau+2}^{T} (\delta R)^{t-\tau-1} u'(\hat{c}_t(a_t)) \hat{c}'_t(a_t)(1 - \hat{c}'_{t-1}(a_{t-1})) \cdots (1 - \hat{c}'_{\tau+2}(a_{\tau+2})).$$

Multiply both sides by $\delta R(1 - \hat{c}'_{\tau+1}(a_{\tau+1}))$ and substitute back in the equation for $u'(\hat{c}_\tau)$

to verify that the statement in the lemma also holds for $t = \tau$:

$$u'(\hat{c}_\tau) = \hat{\beta}(\delta R)u'(\hat{c}_{\tau+1})\hat{c}'_{\tau+1}(a_{\tau+1}) + (\delta R)u'(\hat{c}_{\tau+1})[1 - \hat{c}'_{\tau+1}(a_{\tau+1})].$$

□

Lemma 15. *The first-period consumption $\overline{c}_1$ satisfies:*

$$u'(\overline{c}_1) = \delta R u'(\hat{c}_2)\frac{\beta}{\hat{\beta}}\left[1 - \hat{c}'_2(a_2) + \hat{\beta}\hat{c}'_2(a_2)\right],$$

where $a_2 = R(a_1 - \overline{c}_1)$.

Proof. Similar to the proof of last lemma,

$$u'(\overline{c}_1) = \beta \sum_{t=2}^{T} (\delta R)^{t-1} u'(\hat{c}_t(a_t))\hat{c}'_t(a_t)(1 - \hat{c}'_{t-1}(a_{t-1}))\cdots(1 - \hat{c}'_2(a_2)).$$

The FOC at $t = 2$ gives to

$$u'(\hat{c}_2) = \hat{\beta} \sum_{t=3}^{T} (\delta R)^{t-2} u'(\hat{c}_t(a_t))\hat{c}'_t(a_t)(1 - \hat{c}'_{t-1}(a_{t-1}))\cdots(1 - \hat{c}'_3(a_3)).$$

Multiply by $\delta R(1 - \hat{c}'_2(a_2))$ on both sides, and substitute back to the equation for $u'(\overline{c}_1)$, then we obtain

$$u'(\overline{c}_1) = \beta\delta R u'(\hat{c}_2)\hat{c}'_2(a_2) + \frac{\beta}{\hat{\beta}}\delta R u'(\hat{c}_2)(1 - \hat{c}'_2(a_2)).$$

□

Lemma 16. $(\delta R)^{t-1}u'(\hat{c}_t(a_t)) \geq \frac{\hat{\beta}}{\beta}u'(\overline{c}_1)$ *for all $t > 1$.*

Proof. It is straightforward to see that $\hat{c}'_t(a_t) \in [0, 1], \forall t > 1$. It follows that $1 - \hat{c}'_t(a_t) +$

$\hat{\beta}\hat{c}_t'(a_t) \in [\hat{\beta}, 1]$. From the previous two lemmas, we have

$$(\delta R)^{t-1} u'(\hat{c}_t(a_t)) \leq (\delta R)^t u'(\hat{c}_{t+1}(a_{t+1})), \forall 1 < t < T.$$
$$u'(\overline{c}_1) \leq \frac{\beta}{\hat{\beta}} \delta R u'(\hat{c}_2).$$

It immediately follows that $(\delta R)^{t-1} u'(\hat{c}_t(a_t)) \geq \frac{\hat{\beta}}{\beta} u'(\overline{c}_1), \forall t > 1.$ □

Lemma 17. *The naive agent consumes weakly more than the sophisticated agent in the first period:* $\overline{c}_1 \geq c_1^S$.

Proof. We argue by contradiction. Suppose $\overline{c}_1 < c_1^S$. Then $u'(\overline{c}_1) > u'(c_1^S)$. From the FOC of the sophisticate's problem, we know that

$$u'(c_1^S) = \frac{\beta}{\hat{\beta}} (\delta R)^{t-1} u'(c_t^S).$$

Together with the previous lemma, we obtain

$$\frac{\beta}{\hat{\beta}} (\delta R)^{t-1} u'(\hat{c}_t(a_t)) > \frac{\beta}{\hat{\beta}} (\delta R)^{t-1} u'(c_t^S).$$

Thus, $\hat{c}_t(a_t) < c_t^S$, which is a contradiction because of the zero-profits condition

$$\overline{c}_1 + \sum_{t=2}^{T} \frac{\hat{c}_t(a_t)}{R^{t-1}} = a_1 = \sum_{t=1}^{T} \frac{c_t^S}{R^{t-1}}.$$

□

We are now ready to show the proposition. Let $\overline{\mathbf{c}} = (\overline{c}_1, \cdots, \overline{c}_T)$ denote the naive consumer's equilibrium allocation. Since $\overline{\mathbf{c}}$ also satisfies the zero-profit condition, a revealed-preference argument applied to the sophisticate's program gives:

$$u(c_1^S) + \frac{\beta}{\hat{\beta}} \sum_{t=2}^{T} \delta^{t-1} u(c_t^S) \geq u(\overline{c}_1) + \frac{\beta}{\hat{\beta}} \sum_{t=2}^{T} \delta^{t-1} u(\overline{c}_t). \tag{sA29}$$

The naive consumer's welfare is:

$$\begin{aligned}\sum_{t=1}^{T}\delta^{t-1}u(\bar{c}_t) &= \left(1-\frac{\hat{\beta}}{\beta}\right)u(\bar{c}_1)+\frac{\hat{\beta}}{\beta}u(\bar{c}_1)+\sum_{t=2}^{T}\delta^{t-1}u(\bar{c}_t)\\ &= \left(1-\frac{\hat{\beta}}{\beta}\right)u(\bar{c}_1)+\frac{\hat{\beta}}{\beta}\left[u(\bar{c}_1)+\frac{\beta}{\hat{\beta}}\sum_{t=2}^{T}\delta^{t-1}u(\bar{c}_t)\right]\\ &\le \left(1-\frac{\hat{\beta}}{\beta}\right)u(\bar{c}_1)+\frac{\hat{\beta}}{\beta}\left[u(c_1^S)+\frac{\beta}{\hat{\beta}}\sum_{t=2}^{T}\delta^{t-1}u(c_t^S)\right]\\ &= \left(1-\frac{\hat{\beta}}{\beta}\right)u(\bar{c}_1)+\frac{\hat{\beta}}{\beta}u(c_1^S)+\sum_{t=2}^{T}\delta^{t-1}u(c_t^S)\\ &\le u(c_1^S)+\sum_{t=2}^{T}\delta^{t-1}u(c_t^S),\end{aligned}$$

where equalities come from algebraic manipulation, the first inequality comes from (sA29), and the last inequality comes from the previous lemma $\bar{c}_1 \ge c_1^S$. So the naive consumer's welfare is bounded above by the sophisticate's welfare, which does not converge to the time-consistent consumer's welfare (Proposition 9), establishing the result.

□

Proof of Proposition 9. We argue by contradiction. Suppose instead that

$$\liminf_{T\nearrow+\infty}\left(W_T^C-W_T^S\right)=0,$$

so that there exists a subsequence $\{T_n : n\in\mathbb{N}\}$ with $\lim_{n\nearrow+\infty}\left(W_{T_n}^C-W_{T_n}^S\right)=0$.

Let $c^S=(c_1^S(T),\cdots,c^S(s_T,T))$ denote the equilibrium consumption for the sophisticated agent in the (truncation of the) model with T periods. Let $c^C=(c_1^C(T),\cdots,c^C(s_T,T))$ denote the equilibrium consumption for the time-consistent agent in the (truncation of the) model with T periods. Passing to subsequences, we can assume both limits $\lim_{n\nearrow\infty}c_1^S(T_n)$ and $\lim_{n\nearrow\infty}c_1^C(T_n)$ exist.[34]

[34]That is, there exists a subsequence $\{T_{n_m}\}$ of $\{T_n\}$ such that the limit of $c_1^S(T_{n_m})$ exists. Similarly, consider the sequence $\{c_1^C(T_{n_{m_k}})\}$. Again, pick a subsequence $\{T_{n_{m_o}}\}$ of $\{T_{n_m}\}$ such that the limit $c_1^C(T_{n_{m_k}})$ exists. For notational simplicity, and with no loss of generality, we can replace the original sequence $\{T_n\}$

We first claim that the sophisticate consumes strictly more in the first period than the time-consistent consumer in the limit: $\lim_{n\nearrow\infty} c_1^S(T_n) > \lim_{n\nearrow\infty} c_1^C(T_n)$. Suppose instead that $\lim_{n\nearrow\infty} c_1^S(T_n) \leq \lim_{n\nearrow\infty} c_1^C(T_n)$. The FOCs of the time-consistent consumer's program give:

$$u'(c_1^C(T_n)) = (\delta R)^{t-1} Eu'(c^C(s_t, T_n)), \quad u'(c_1^S(T_n)) = \beta(\delta R)^{t-1} Eu'(c^S(s_t, T_n)).$$

We claim that $\liminf_{n\nearrow\infty}(c^S(s_t, T_n) - c^C(s_t, T_n)) < 0, \forall t > 1$. Otherwise, there exists $t > 1$ and $\liminf_{n\nearrow\infty}(c^S(s_t, T_n) - c^C(s_t, T_n)) \geq 0$. Passing to subsequences, we can assume that $\lim_{n\nearrow\infty} c_t^S(s_t, T_n)$ and $\lim_{n\nearrow\infty} c_t^C(s_t, T_n)$ exist.

It follows that

$$\begin{aligned}
\lim_{n\nearrow\infty} (\delta R)^{t-1} Eu'(c^S(s_t, T_n)) &> \lim_{n\nearrow\infty} \beta(\delta R)^{t-1} Eu'(c^S(s_t, T_n)) \\
&= \lim_{n\nearrow\infty} u'(c_1^S(T_n)) \\
&\geq \lim_{n\nearrow\infty} u'(c_1^C(T_n)) \\
&= \lim_{n\nearrow\infty} (\delta R)^{t-1} Eu'(c^C(s_t, T_n)) \\
&\geq \lim_{n\nearrow\infty} (\delta R)^{t-1} Eu'(c^S(s_t, T_n)),
\end{aligned}$$

where the first inequality is strict because of $\beta < 1$, the second equation comes from the sophisticate's FOC, the third comes from $\lim_{n\nearrow\infty} c_1^S(T_n) \leq \lim_{n\nearrow\infty} c_1^C(T_n)$, the fourth comes from the time-consistent consumer's FOC, the last comes from $\liminf_{n\nearrow\infty}(c^S(s_t, T_n) - c^C(s_t, T_n)) \geq 0$. This is a contradiction. So $\liminf_{n\nearrow\infty}(c^S(s_t, T_n) - c^C(s_t, T_n)) < 0, \forall t > 1$. But then it violates the zero-profit condition since

$$0 = \liminf_{n\nearrow\infty} \sum_{t=1}^{T_n} \frac{E(c^S(s_t, T_n) - c^C(s_t, T_n))}{R^{t-1}} < 0.$$

What we have shown now is that in the first period the sophisticate consume strictly more

with this last subsequence $\{T_{n_{m_o}}\}$.

than the time consistent consumer in the limit: $\lim_{n\nearrow\infty} c_1^S(T_n) > \lim_{n\nearrow\infty} c_1^C(T_n)$.

Define $l_T(c_1)$ as

$$l_T(c_1) = \max_{c(s_2),\cdots,c(s_T)} \sum_{t=2}^{T} \delta^{t-1} u(c(s_t)),$$

subject to $E\sum_{t=2}^{T} \frac{c(s_t)}{R^{t-1}} = E\sum_{t=1}^{T} \frac{w(s_t)}{R^{t-1}} - c_1$. We claim that $l_T''(c_1) < 0$. Let λ_l denote the Lagrangian on the zero-profit constraint.

$$l_T(c_1) = \sum_{t=2}^{T} \delta^{t-1} u(c(s_t)) + \lambda_l \left(E\sum_{t=1}^{T} \frac{w(s_t)}{R^{t-1}} - c_1 - E\sum_{t=2}^{T} \frac{c(s_t)}{R^{t-1}} \right).$$

Taking derivative with respect to c_1: $l_T'(c_1) = -\lambda_l$. Then, $l_T''(c_1) = -\lambda_l'$.

Taking derivative on the both sides of FOC, $\delta^{t-1}u'(c_t) = \frac{\lambda_l}{R^{t-1}}$, with respect to c_1:

$$\frac{\partial c_t}{\partial c_1} = \frac{\lambda_l'}{(\delta R)^{t-1} u''(c_t)}.$$

Taking derivative with respect to c_1 on the zero-profit condition:

$$-1 = \sum_{t=2}^{T} \frac{\frac{\partial c_t}{\partial c_1}}{R^{t-1}} = \sum_{t=2}^{T} \frac{\lambda_l'}{(\delta R)^{t-1} u''(c_t)}.$$

Thus, $\lambda_l' > 0$ because $u'' < 0$. So $l_T''(c_1) = -\lambda_l' < 0$.

It implies that $u(c_1) + l_T(c_1)$ is a concave function of c_1 for any T. Since u is bounded and $\delta < 1$, we can use dominated convergence theorem. Taking limit of T to infinity, $\limsup_{T\nearrow\infty} u''(c_1) + l_T''(c_1) \leq \limsup_{T\nearrow\infty} u''(c_1) < 0$, since we assume strict concavity of u. So $\lim_{T\nearrow\infty}[u(c_1) + l_T(c_1)]$ is a strict concave function of c_1. Together with our first claim that $\lim_{n\nearrow\infty}(c_1^S(T_n) - c_1^C(T_n)) > 0$ and the fact that $c_1^C(T)$ maximizes $u(c_1) + l_T(c_1)$, it follows that

$$\lim_{n\nearrow+\infty} \left(u(c_1^C(T_n)) + l_T(c_1^C(T_n)) - u(c_1^S(T_n)) - l_T(c_1^S(T_n)) \right) > 0,$$

i.e. $\lim_{n\nearrow+\infty} \left(W_{T_n}^C - W_{T_n}^S \right) > 0$, a contradiction. So the welfare loss for sophisticated agents is bounded below away from 0. □

Proof of Lemma 9. We first show that the (IC) constraints for self-2 must be binding for all $m_2 \in supp(\sigma_2)$ and $m_2' \in supp(\hat{\sigma}_2)$. We note that the (IC) must be binding for at least one m_2', because otherwise we can increase consumption on the perceived path and increase the self 1's payoff. Now suppose there exists $m_2 \in supp(\sigma_2)$ and $m_2'' \in supp(\hat{\sigma}_2)$ such that the corresponding (IC) is slack. In this case, we show that from self 1's perspective, the perceived path m_2'' gives a higher payoff than the perceived path m_2' (using a coefficient of 1). To see that, notice that from (PC) constraint, the perceived self-2 is indifferent between m_2' and m_2'' (using a coefficient of $\hat{\beta}$), but self-2 strictly prefers m_2' over m_2'' (using a coefficient of β). By the single crossing property, it implies that m_2'' gives a strictly higher payoff than m_2' in calculating self 1's perceived payoff (using a coefficient of 1). Then, replacing terms in options m_2' with terms in options m_2'' would not affect any constraints, but it would increase self 1's perceived payoff, a contradiction to the optimality of the original contract.

Next we show that $c_2(s_2, \hat{m}_2) = 0, \forall \hat{m}_2 \in supp(\hat{\sigma}_2)$. Otherwise, consider a perturbation in which lowers $u(c_2(s_2, \hat{m}_2))$ by $\beta\epsilon$ and increases $u(c_T(s_T, \hat{m}_2, \hat{\sigma}_3, \cdots, \hat{\sigma}_{T-1}))$ by ϵ. This perturbation preserves the IC constraints and maintains all other constraints, but increases self 1's perceived payoff.

Substituting the binding IC constraint into the objective function, we obtain (up to a constant):

$$\max_{\{c(s_t,h^t)\}} u(c(s_1)) + \delta E u(c(s_2, \sigma_2)) + \beta E\left[\sum_{t=3}^{T} \delta^{t-1} u(c(s_t, \sigma_2, \hat{\sigma}_3, ..., \hat{\sigma}_{T-1}))\right].$$

Repeating the same analysis, we have a new program (up to a constant):

$$\max_{\{c(s_t,h^t)\}} E\delta^{t-1} u(c(s_t, \sigma_2, \cdots, \sigma_{T-1})) + \beta E\left[\delta^{t-1} u(c(s_T, \sigma_2, \sigma_3, ..., \sigma_{T-1}))\right].$$

subject to the zero-profit condition.

Our final step is showing that the equilibrium path $\sigma_2, \sigma_3, ..., \sigma_{T-1}$ involves only one option. This is because of Jensen's inequality and the strict concavity of $u(\cdot)$. If there are multiple options in the σ_τ, then merging those options can strictly increase self 1's perceived payoff. This completes the proof. $\square$

Proof of Lemma 2 for General Income Distributions and Arbitrary T

This appendix establishes the equivalence between the naive agent's program and the auxiliary program for general income distributions and arbitrary T. As in the text, we consider the one-sided commitment case. With two-sided commitment, one can ignore the non-lapsing constraints in the proof below.

Recall that the naive agent's program is

$$\max_{c(s_t,h^t)} u(c(s_1)) + \beta E\left[\sum_{t=2}^{T} \delta^{t-1} u(c(s_t, (B, B, ..., B)))\right],$$

subject to

$$\sum_{t=1}^{T} E\left[\frac{w(s_t) - c(s_t, (A, A, ..., A))}{R^{t-1}}\right] = 0, \qquad \text{(Zero Profits)}$$

$$u(c(s_\tau, (h^{\tau-1}, B))) + \hat{\beta} E\left[\sum_{t>\tau} \delta^{t-\tau} u\left(c\left(s_t, \left(h^{\tau-1}, B, B, ..., B\right)\right)\right) \middle| \, s_\tau\right] \qquad \text{(PCC)}$$

$$\geq u(c(s_\tau, (h^{\tau-1}, A))) + \hat{\beta} E\left[\sum_{t>\tau} \delta^{t-\tau} u\left(c\left(s_t, \left(h^{\tau-1}, A, B, ..., B\right)\right)\right) \middle| \, s_\tau\right],$$

and

$$u(c(s_\tau, (h^{\tau-1}, A))) + \beta E\left[\sum_{t>\tau} \delta^{t-\tau} u\left(c\left(s_t, \left(h^{\tau-1}, A, B, ..., B\right)\right)\right) \middle| \, s_\tau\right] \qquad \text{(IC)}$$

$$\geq u(c(s_\tau, (h^{\tau-1}, B))) + \beta E\left[\sum_{t>\tau} \delta^{t-\tau} u\left(c\left(s_t, \left(h^{\tau-1}, B, B, ..., B\right)\right)\right) \middle| \, s_\tau\right],$$

and non-lapsing constraints:

$$u\left(c(s_\tau,(h^{\tau-1},A))\right)+\beta E\left[\sum_{t>\tau}\delta^{t-\tau}u\left(c(s_t,(h^{\tau-1},A,B,...,B))\right)\middle| s_\tau\right]\geq V(s_\tau),\ \ \forall s_\tau, \tag{NL}$$

and

$$u\left(c(s_\tau,(h^{\tau-1},B))\right)+\hat{\beta} E\left[\sum_{t>\tau}\delta^{t-\tau}u\left(c(s_t,(h^{\tau-1},B,B,...,B))\right)\middle| s_\tau\right]\geq \hat{V}(s_\tau),\ \ \forall s_\tau. \tag{PNL}$$

We first note that the incentive compatibility constraints (IC) must be binding on the equilibrium path, because otherwise we can increase $c(s_T,h^\tau,B,B,...,B)$ without affecting all other constraints while weakly increase the agent's perceived utility. Given incentive constraints are binding, we can simplify (PC) as

$$u(c(s_\tau,(h^{\tau-1},B)))\leq u(c(s_\tau,(h^{\tau-1},A))). \tag{sA30}$$

Substituting the binding IC constraints in the objective gives

$$E\sum_{t=1}^{T-1}\delta^{t-1}u(c(s_t,A,\cdots,A))+\beta\delta^{T-1}u(c(s_T,A,\cdots,A))+(\beta-1)\delta^{t-1}u(c(s_t,A,\cdots,A,B)).$$

Since $\beta<1$, we want to choose $c(s_t,A,\cdots,A,B)$ as small as possible (subject to the constraints). We now show that under the optimal contract, $c(s_t,A,\cdots,A,B)=0$. We need to verify that setting $c(s_t,A,\cdots,A,B)=0$ would not violate all other constraints. First, PC holds because (sA30) holds.

We then verify that PNL holds if NL holds. Suppose $\{\hat{c}(s_t,h^t_\tau):t\geq\tau\}$ solves the perceived outside option program $\hat{V}^I(s_\tau)$. So we have

$$\hat{V}^I(s_\tau)=u(\hat{c}(s_\tau,h^\tau_\tau))+\hat{\beta}E[\delta^{t-\tau}u(\hat{c}(s_t,(h^\tau_\tau,B,\cdots,B)))|s_\tau]. \tag{sA31}$$

We next verify the perceived non-lapsing constraint at $(s_\tau,(h^{\tau-1},B))=(s_\tau,(A,\cdots,A,B))$.

Other perceived non-lapsing constraints can be similarly verified. Note that

$$u(c(s_\tau,(h^{\tau-1},B)))+\hat{\beta}E\left[\sum_{t>\tau}\delta^{t-\tau}u(c(s_t,(h^{\tau-1},B,\cdots,B)))\middle|s_\tau\right]$$

$$=\left(1-\frac{\hat{\beta}}{\beta}\right)u(0))+\frac{\hat{\beta}}{\beta}u(c(s_\tau,(h^{\tau-1},B)))+\hat{\beta}E\left[\sum_{t>\tau}\delta^{t-\tau}u(c(s_t,(h^{\tau-1},B,\cdots,B)))\middle|s_\tau\right] \tag{sA32}$$

$$=\left(1-\frac{\hat{\beta}}{\beta}\right)u(0)+\frac{\hat{\beta}}{\beta}\left(u(c(s_\tau,(h^{\tau-1},A)))+\beta E\left[\sum_{t>\tau}\delta^{t-\tau}u(c(s_t,(h^{\tau-1},A,B,\cdots,B)))\middle|s_\tau\right]\right) \tag{sA33}$$

$$\geq\left(1-\frac{\hat{\beta}}{\beta}\right)u(0)+\frac{\hat{\beta}}{\beta}V^I(s_\tau) \tag{sA34}$$

$$\geq\left(1-\frac{\hat{\beta}}{\beta}\right)u(0)+\frac{\hat{\beta}}{\beta}\left(u(\hat{c}(s_\tau,h_\tau^\tau))+\beta E\left[\sum_{t>\tau}\delta^{t-\tau}u(\hat{c}(s_t,h_\tau^\tau,B,\cdots,B))\middle|s_\tau\right]\right) \tag{sA35}$$

$$=\left(1-\frac{\hat{\beta}}{\beta}\right)u(0)+\frac{\hat{\beta}}{\beta}u(\hat{c}(s_\tau,h_\tau^\tau))+\hat{\beta}E\left[\sum_{t>\tau}\delta^{t-\tau}u(\hat{c}(s_t,h_\tau^\tau,B,\cdots,B))\middle|s_\tau\right] \tag{sA36}$$

$$=\left(1-\frac{\hat{\beta}}{\beta}\right)u(0)+(\frac{\hat{\beta}}{\beta}-1)u(\hat{c}(s_\tau,h_\tau^\tau))+\hat{V}^I(s_\tau) \tag{sA37}$$

$$\geq\hat{V}^I(s_\tau), \tag{sA38}$$

where (sA32) follows from $c(s_\tau,(h^{\tau-1},B))=0$, (sA33) from (IC), (sA34) from the actual non-lapsing constraints (NL), (sA35) follows from a revealed preference argument since $\hat{c}$ is also feasible in program $V(s_\tau)$, (sA36) follows from simple algebra, (sA37) uses (sA31), and (sA38) follows from $\hat{c}(s_\tau,h_\tau^\tau)\geq 0$.

By the previous argument, the perceived choice constraints and the perceived non-lapsing constraints can be ignored, so the program reduces to:

$$\max E\sum_{t=1}^{T-1}\delta^{t-1}u(c(s_t,(A,\cdots,A)))+\beta\delta^{T-1}u(c(s_T,(A,\cdots,A))),$$

subject to the zero-profit condition and the non-lapsing constraints. Since the objective is the same as the utility of a dynamically consistent consumer, we can replace the non-lapsing constraints by front-loading constraints. So $c^{1E} = c^{1A}$. $\square$

Corollary 1 with One-Sided Commitment

This appendix generalizes Corollary 1 for settings with one-sided commitment, as mentioned in footnote 18:

Corollary 2. *Consider the model with one-sided commitment. There exists a perception-perfect equilibrium that does not depend on the consumer's naiveté $\hat{\beta} \in (\beta, 1]$. Moreover, any perception-perfect equilibrium has the same consumption path, which is continuous in $\beta \in (0, 1]$.*

Proof. By Lemma 4, we can focus on the auxiliary program with one-sided commitment. Let $x(s_t) \equiv u(c(s_t))$ denote the agent's utility from the consumption he gets in state s_t, and consider the dual program:

$$\max_{\{x(s_t)\}} \sum_{t=1}^{T} \sum_{s_t \in \mathbb{S}_t} p(s_t|s_1) \frac{w(s_t) - u^{-1}(x(s_t))}{R^{t-1}}, \tag{sC1}$$

subject to

$$\sum_{t=1}^{T-1} \sum_{s_t \in \mathbb{S}_t} \delta^{t-1} p(s_t|s_1) x(s_t) + \beta \sum_{s_T \in \mathbb{S}_T} \delta^{T-1} p(s_T|s_1) x(s_T) \geq \underline{u}, \tag{sC2}$$

and

$$\sum_{t \geq \tilde{\tau}} \sum_{s_t \in \mathbb{S}_t} \delta^{t-\tilde{\tau}} p(s_t|s_{\tilde{\tau}}) x(s_t) + \beta \sum_{s_T \in \mathbb{S}_T} \delta^{T-\tilde{\tau}} p(s_T|s_{\tilde{\tau}}) x(s_T) \geq V_T^A(s_{\tilde{\tau}}) \;\; \forall s_{\tilde{\tau}} \in \mathbb{S}_{\tilde{\tau}}(s_\tau), \forall \tau, \tag{sC3}$$

This program corresponds to the maximization of a strictly concave function over a convex set, so, by the Theorem of the Maximum, the solution is unique and continuous in $\beta \in$

$(0, 1]$. Moreover, since the program does not involve $\hat{\beta}$, the equilibrium consumption path is not a function of the consumer's naiveté.

Once we pin down the unique consumption path, the baseline options are either zero or determined by the binding IC constraints and non-lapsing constraints (See the proof of Lemma 4). In particular, these constraints do not depend on the consumer's naiveté. So the equilibrium consumption vector is not a function of the consumer's naiveté. □

Removing Commitment Power

This appendix presents the formal analysis of the welfare effect of removing commitment power, as described in Subsection 3.1. We show that, for a fixed contract length, removing commitment power can make the consumer better off. To formalize the argument given in the text, let $\mathcal{V}_T^S$ denote the agent's welfare from smoothing consumption perfectly in the first $T-1$ periods and consuming zero in the last period:

$$\mathcal{V}_T^S \equiv \max_{\{c(s_t)\}} \sum_{t=1}^{T-1} E\left[\delta^{t-1} u\left(c(s_t)\right)\right] + \delta^{T-1} u(0),$$

subject to

$$\sum_{t=1}^{T-1} E\left[\frac{c(s_t)}{R^{t-1}}\right] \leq \sum_{t=1}^{T} E\left[\frac{w(s_t)}{R^{t-1}}\right].$$

Let $\mathcal{V}_T^{NS}$ denote the agent's welfare from consuming the endowment in each state:

$$\mathcal{V}_T^{NS} \equiv \sum_{t=1}^{T} E\left[\delta^{t-1} u\left(w(s_t)\right)\right].$$

Proposition 10. *Suppose agents are time inconsistent and $\mathcal{V}_T^{NS} > \mathcal{V}_T^S$. There exists $\bar{\beta} > 0$ such that if $\beta < \bar{\beta}$, the welfare with one-sided commitment is greater than the welfare with two-sided commitment.*

Proof. First, note that the welfare with two-sided commitment approaches to $\mathcal{V}_S$ as β approaches to zero. It suffices to show that the welfare with one-sided commitment is bounded below by $\mathcal{V}_{NS}$. In the remainder of the proof, we will therefore focus on the equilibrium with one-sided commitment.

We claim that for β close to zero, the equilibrium consumption equals the endowment in all last-period states: $c(s_T) = w(s_T), \forall s_T \in \mathbb{S}_T(s_1)$. To see this, consider a perturbation that shifts consumption from a state in the last period to the preceding state, that is, it increases $c(s_{T-1})$ by $\epsilon > 0$ and reduces $c(s_T)$ by $\frac{\epsilon R}{p(s_T|s_{T-1})}$ for some $s_T \in \mathbb{S}_T$ with $p(s_T|s_{T-1}) > 0$. Let W_{s_T} denote the future value of all income up to state s_T. The amount W_{s_T} is how much the agent would be able to consume at state s_T if he saves all his income from all periods for the last one. It therefore gives an upper bound on how much the agent can consume in the last period. Since there are finitely many states and $W_{s_T} < \infty$ for all s_T, we can take the uniform bound $W \equiv \max_{s_T} W_{s_T}$. This perturbation affects the LHS of the non-lapsing constraint at state s_t by

$$\begin{aligned} &p(s_{T-1}|s_t)\left[u'(c(s_{T-1})) - \beta R\delta u'(c(s_T))\right]\delta^{T-1-t}\epsilon \\ &> p(s_{T-1}|s_t)\left[u'(0) - \beta R\delta u'(W_{s_T})\right]\delta^{T-1-t}\epsilon, \end{aligned}$$

which is positive whenever

$$\frac{u'(0)}{R\delta u'(W)} > \beta. \tag{sC1}$$

The perturbation has exactly the same effect on the objective function (scaled down by δ^t and multiplied by the probability of reaching state s_{T-1}). Thus, as long as β satisfies (sC1), the equilibrium will have the smallest consumption possible in the last period, which is determined by the non-lapsing constraint.

Substituting $c(s_T) = w(s_T)$ in the auxiliary program, it becomes analogous to the program of a time-consistent agent except that the contracting problem ends at period $T-1$ instead of period T:

$$\max_{\{c(s_t)\}} \sum_{t=1}^{T-1} \sum_{s_t \in \mathbb{S}_t(s_1)} \delta^{t-1} p(s_t|s_1) u\left(c(s_t)\right),$$

subject to

$$\sum_{t=1}^{T-1} \sum_{s_t \in \mathbb{S}_t(s_1)} p(s_t|s_1) \frac{w(s_t) - c(s_t)}{R^{t-1}} = 0,$$

and

$$\sum_{t=\tilde{\tau}}^{T-1} \sum_{s_t \in \mathbb{S}_t(s_{\tilde{\tau}})} \delta^{t-\tilde{\tau}} p(s_t|s_{\tilde{\tau}}) u\left(c(s_t)\right) \geq (V')^C(s_{\tilde{\tau}}) \text{ for all } s_{\tilde{\tau}},$$

for all $\tilde{\tau} = 2, ..., T$, where $(V')^C(s_{\tilde{\tau}})$ denotes the outside option for the time-consistent agent in this $(T-1)$-period economy.

It is straightforward to verify that $(V')^C(s_1)$ is bounded below by the utility from consuming the endowment in all states. If the endowment already satisfies the non-lapsing constraints, then the result follows from revealed preference because the endowment also satisfies zero profits. If the endowment does not satisfy the non-lapsing constraints, any renegotiation of the endowment satisfies the zero-profits condition and gives the time-consistent agent a strictly higher utility conditional on that state. So, replacing the endowment by the solution of the continuation program in all states where the non-lapsing constraints are violated leads to a profile of consumption that satisfies the constraints and gives a utility greater than the utility of consuming the endowment in each period. It thus follows by revealed preference that the solution of the program also gives a higher utility than consuming the endowment in all states.

Since the solution of a naive agent coincides with the solution of this auxiliary program, their welfare is also bounded below by the welfare from consuming their endowment in all periods $\mathcal{V}_{NS}$ when (sC1) holds. Therefore, by continuity, if $\mathcal{V}_{NS} > \mathcal{V}_S$, there exists $\bar{\beta}_N$ such that if $\beta < \bar{\beta}_N$, the welfare with one-sided commitment dominates the welfare with two-sided commitment. □

Notice that for generic endowment paths, the condition that $\mathcal{V}_T^{NS} > \mathcal{V}_T^S$ fails when T is large enough. So, as the contracting length grows, it becomes increasingly hard to satisfy the conditions for the time-inconsistent consumer to obtain higher welfare without commitment, as described in the text.

Proof of Claim in Section 3.5

In this appendix, we establish that a naive agent saves more than a sophisticate in the first period. Given a vector $x = (x_1, x_2, \cdots, x_T)$ with $x_1 = 1$ consider the program:

$$\max_{(c_1,\cdots,c_T)} \sum_{t=1}^{T} x_t u(c_t), \tag{sD1}$$

subject to

$$\sum_{t=1}^{T} \frac{c_t}{R^{t-1}} = \sum_{t=1}^{T} \frac{w}{R^{t-1}}. \tag{sD2}$$

The first-order conditions of this program are:

$$R^{t-1} x_t u'(c_t) \leq \lambda, \ \ \forall t, \tag{sD3}$$

where λ is the Lagrangian multiplier on the zero-profits condition (sD2). The inequality becomes equality whenever $c_t > 0$.

The consumption path of a naive agent solves this program for

$$x^N = \left(1, \frac{D_{T-1}}{D_{T-2}}, \frac{D_{T-1}}{D_{T-3}}, \cdots, \frac{D_{T-1}}{D_1}, D_{T-1}\right),$$

whereas the consumption path of a sophisticated agent solves this program for vector

$$x^S = (1, D_1, D_2, \cdots, D_{T-1}).$$

Let λ^N and λ^S denote the Lagrangian multiplier in the case of a naive agent and a sophisticated agent, respectively. Recall from equation (29), $x_t^N \geq x_t^S$ for all $t = 2, \cdots, T$.

We argue by contradiction and suppose the naive agent consumes strictly more than the sophisticate in the first period, i.e., $c_1^N > c_1^S$. We claim that $c_t^N \geq c_t^S, \forall t$. Then the claim together with $c_1^N > c_1^S$ would violate the zero-profits condition.

To prove the claim, first note that the claim trivially holds if $c_t^S = 0$. Now suppose $c_t^S > 0$, then $R^{t-1} x_t^S u'(c_t^S) = \lambda^S \geq u'(c_1^S)$. Since $c_1^N > c_1^S \geq 0$, the FOC at c_1^N is an equality: $u'(c_1^N) = \lambda^N$.

Note that for any $t = 2, \cdots, T$,

$$R^{t-1} x_t^S u'(c_t^S) = \lambda^S \geq u'(c_1^S) \geq u'(c_1^N) = \lambda^N \geq R^{t-1} x_t^N u'(c_t^N) \geq R^{t-1} x_t^S u'(c_t^N),$$

where the last inequality comes from $x_t^N \geq x_t^S$. It follows that $u'(c_t^S) \geq u'(c_t^N)$, i.e., $c_t^N \geq c_t^S, \forall t = 2, \cdots, T$. Together with $c_1^N > c_1^S$, it contradicts to the zero-profits condition. As a result, the naive agent must consume weakly less than the sophisticate in the first period (i.e., the naive agent saves weakly more than the sophisticate in the first period).

Moreover, the naive agent must consume *strictly* less than the sophisticate in the first period if $\lim_{c \searrow 0} u'(c) = +\infty$. In this case, we have an interior solution, and (sD3) becomes equality because consumption is always strictly positive. To see that $c_1^N < c_1^S$, we need to show that there is a contradiction when $c_1^N = c_1^S$. We recall that $x_t^N > x_t^S$ for all $t = 2, \cdots, T-1$. Now for $t = 2, \cdots T-1$,

$$R^{t-1} x_t^S u'(c_t^S) = u'(c_1^S) = u'(c_1^N) = R^{t-1} x_t^N u'(c_t^N) > R^{t-1} x_t^S u'(c_t^N),$$

which implies that $c_t^N > c_t^S$, for $2 \leq t \leq T-1$. We still have $c_T^N \geq c_T^S$. Together, we have a contradiction to the zero-profits condition. So if $\lim_{c \searrow 0} u'(c) = +\infty$, the naive agent must consume *strictly* less than the sophisticated agent in the first period.