1-1-2020

# The Welfare Effects of Long-Term Health Insurance Contracts

Soheil Ghili

Ben Handel

Igal Hendel

Michael D. Whinston

# THE WELFARE EFFECTS OF LONG-TERM HEALTH INSURANCE CONTRACTS

By

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

December 2019

COWLES FOUNDATION DISCUSSION PAPER NO. 2218

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

# The Welfare Effects of Long-Term Health Insurance Contracts*

Soheil Ghili[†], Ben Handel[‡], Igal Hendel[§], and Michael D. Whinston[¶]

January 2017
(This Draft: December 2019)

**Abstract**

Reclassification risk is a major concern in health insurance where contracts are typically one year in length but health shocks often persist for much longer. We use rich individual-level medical information from the Utah all-payer claims database to empirically study one possible solution: long-term insurance contracts. We characterize optimal long-term contracts with one-sided commitment theoretically, derive the contracts that are optimal for consumers in Utah, and assess the welfare level that a full implementation of these contracts could achieve relative to several key benchmarks. We find that dynamic contracts perform very well for the majority of the population, for example, eliminating over 94% of the welfare loss from reclassification risk for individuals who arrive on the market at age 25 in good health. However, dynamic contracts instead provide very little benefit to the worst pre-age-25 health risks. Their value is also substantially lower for consumers whose income growth with age is relatively high. With pre-age-25 insurance in place, consumers with flat net income prefer dynamic contracts to an ACA-like environment, but consumers with steeper income profiles prefer the ACA-like environment. Overall, we show that there are scenarios in which

* We thank Nathan Hendren, Kurt Lavetti, and Neale Mahoney, as well as seminar participants at Arizona, Bocconi, EIEF, Georgetown, Harvard, the University of Minnesota Macroeconomics of Pensions Conference, the MIT Public Finance lunch and internal faculty seminar, the April 2017 NBER joint public finance/insurance conference, Penn, Princeton, Stanford, Toulouse, UCLA, UNC, Washington University, and Yale for their comments. All authors are grateful for support from NSF grant SES-1259770. We are grateful to Kurt Lavetti for his guidance with respect to the Utah all-payers claims data. We thank Michael Sullivan for outstanding research assistance.

[†]Yale School of Management; soheil.ghili@yale.edu

[‡]Department of Economics, UC Berkeley; handel@berkeley.edu

[§]Department of Economics, Northwestern University; igal@northwestern.edu

[¶]Department of Economics and Sloan School of Management, M.I.T.; whinston@mit.edu

dynamic contracts can provide substantial welfare benefits, but that complementary policies are crucial for unlocking these benefits.

# 1 Introduction

Consumers face substantial health risks over their lifetimes. Much of this risk involves conditions, such as diabetes, heart disease, and cancer, that lead to high expected medical expenses over significant periods of time. These conditions can expose individuals who buy short-term insurance coverage to substantial premium increases – so-called "reclassification risk" – greatly reducing the extent to which their health risks are insured.

Concerns over reclassification risk have received a great deal of public and academic attention in recent years. In the U.S., the Affordable Care Act (ACA) was designed to contend with this problem through community rating and guaranteed issuance, thereby prohibiting discrimination against consumers who have developed pre-existing conditions. Unfortunately, while requiring identical pricing for consumers with different health can eliminate reclassification risk, it can also create adverse selection, leading to under-provision of insurance [Handel, Hendel and Whinston (2015), Patel and Pauly (2002)].

In this paper we explore an alternative approach: long-term contracts without pricing regulation. By specifying long-term obligations, such contracts can mitigate reclassification risk without causing the adverse selection that forced community rating creates. However, while their appeal is clear, the actual potential benefits from long-term contracts is an empirical matter.

We develop a dynamic model to characterize optimal dynamic health insurance contracts and then implement that model empirically using rich all-payer claims data from the state of Utah to assess the extent of reclassification risk that these contracts eliminate. We compare dynamic contracts to a range of alternative regulatory environments, including, e.g., the ACA, and study the welfare implications of a counterfactual system-wide switch to dynamic contracts for Utah's consumers. We show that dynamic contracts can provide substantial benefits to consumers in some cases, but to be effective would likely require a set of complementary policies and regulations. We also present an extended discussion of these issues, focusing on (i) why such contracts are not common in practice in health insurance markets (as they are in life insurance markets) and (ii) what complementary regulations (such as the elimination of the employer insurance tax-advantage, and provision of government insurance of early-life health risk) could be enacted to increase their prevalence.

We begin our analysis by characterizing optimal long-term health insurance contracts theoretically. If only spot contracts are available, risk-averse consumers are able to fully insure within-period risk but premiums reflect the information revealed over time, implying that consumers fully bear the risk of persistent health shocks. If both consumers and firms could commit to long-term contracts ex ante, prior to information revelation, then the efficient (first-best) allocation of full long-run insurance is possible. Since it is generally acknowledged that consumers cannot be bound to long-term insurance contracts, we focus on contracts with one-sided commitment, in which firms but not consumers can make long-term contractual promises.[1] We model the insurance contracting problem as one of symmetric learning as in Harris and Holmstrom (1983), who study labor markets, and Hendel and Lizzeri (2003), who study life insurance markets. In the model, consumers seek to insure against negative health shocks over their lifetimes. Consumers and insurers learn about these shocks symmetrically. We assume that there are capital market imperfections that prevent consumers from borrowing large amounts due, for example, to a lack of collateral.[2]

As shown in Harris and Holmstrom (1983), the competitive equilibrium in markets with one-sided commitment only partially insures reclassification risk. We show theoretically that optimal unilateral commitment contracts offer consumers a minimum guaranteed consumption level over time, as a function of their risk preferences, income profiles, and stochastic health state transition processes. Dynamically, this minimal guarantee is bumped up to a new level when a consumer has unexpectedly good health realizations (which can involve remaining healthy as the consumer ages), in order to match competitive offers from other insurers and ensure that consumers won't leave the long-term contract. These consumption floors are the counterpart of the downwardly rigid wages in Harris and Holmstrom (1983) and, in our context, imply that optimal contracts involve "front-loading" (premiums in excess of expected medical costs) to lock consumers into the contract and allow insurers to both break even and offer insurance against reclassification risk.

We also show that these optimal contracts can be equivalently offered as guaranteed premium path contracts, which simply promise a fixed path of annual premiums, and that, when they are offered in this form, they are self-selective in the sense that consumers with different lifetime income profiles and risk preferences will choose the contracts designed for them from a menu of optimal contracts, even if firms do not observe this information about individual consumers.[3]

[1] See, for example, the discussions in Diamond (1992), Cochrane (1995), and Pauly, Kunreuther, and Hirth (1995).

[2] Were consumers able to borrow freely, the first best could be achieved by having consumers pay all premiums up front. In practice, these capital market imperfections likely stem from similar factors to those that prevent consumers from committing to make large ex post payments to an insurer.

[3] This form of contract is the counterpart to widely used "Annual Renewable Term" contracts in the life insurance market (see Hendel and Lizzeri (2003)). They are also observed in the German health insurance market (Browne and Hoffmann (2013)).

Our model of long-term contracts relates closely to several prior models, most notably Pauly, Kunreuther, and Hirth (1995) and Cochrane (1995). Pauly, Kunreuther, and Hirth (1995) focus on "guaranteed renewable" contracts that ensure that an insured can renew future coverage at the same rates that the healthiest possible type would pay, while Cochrane (1995) proposes the use of "premium insurance" as a means of insuring against long-term negative shocks to health. We discuss the relative features of our model in depth, and show that these two prior approaches (i) are equivalent to each other and (ii) achieve welfare that can be significantly lower in our empirical context than with our optimal dynamic contracts.

We implement our model empirically with granular data from the Utah all-payer claims database, which contains individual-level diagnostic and spending information on all privately insured consumers in the state of Utah (~2 million consumers). This is also a policy-relevant environment, since states have increasingly gained regulatory responsibility for health insurance markets in recent years (commensurate with their authority prior to the passage of the ACA). For each consumer and each year they are in the data, we use their medical diagnostic codes and spending data to estimate a health index that predicts their medical spending for the upcoming year. With these estimates in hand, we estimate a second-order Markov process that captures the empirical evolution of consumer health. We estimate this Markov process as a function of age and use these estimates to study health risk over the life-cycle, a key ingredient for assessing the welfare implications of dynamic insurance contracts.

We combine our estimates with our model to compute optimal dynamic contracts for male consumers in Utah. We find that a healthy 25 year old male in Utah with a flat net income profile over time ("net income" is income minus expected medical costs), pays a premium of $2,294 despite expected costs of only $837 in that year. The extent of contract front-loading is inverted U-shaped in health status: it is highest for individuals in medium health states and lower for the healthiest and sickest consumers. This is because the extent of front-loading at any point in time depends on both a consumer's current health state and the implications of that current state for future states. Consumers in good health can better afford the front-loading but may benefit less from it, while those in worse health benefit more from the future promises that front-loading buys but may have a hard time affording it. In addition to health status, we show that consumers' expected income profiles play a crucial role in shaping optimal dynamic contracts. When consumers have steeper increasing income profiles, they are asked to front-load less in the optimal contract because their marginal value of incremental income is relatively higher when they are younger. Thus, their income growth over time limits their desire to insure future health risks with valuable current income.

With these optimal dynamic contracts in hand, we study the welfare implications of a full implementation of these contracts, relative to alternative insurance system designs. We do this using a life cycle model of consumer utility that accounts for long-run risk, short-run risk, risk aversion, and expected income profiles.

To better understand the baseline level of potential reclassification risk in our environment, we first investigate the welfare loss from a market with only year-to-year spot contracts (and no community rating), relative to the first best. Across a range of income profiles, spot contracts result in a lifetime welfare loss between 17.8% and 40.6% for consumers who start their adult lives in good health. Since, for some income paths, the welfare loss induced by spot contracts comes in part from their inability to smooth consumption over time, we consider as well an alternative benchmark: contracts that offer full insurance but with premiums at each age equal to the average medical expenses of that age group ("No Borrowing / No Savings Full Insurance"). Relative to this alternative benchmark, spot contracts still entail a 9.3% to 17.8% welfare loss for consumers who start in good health, representing the loss from not being able to insure reclassification risk.

Turning to dynamic contracts, we first investigate their implications for consumers conditional on their age-25 health state. We find that, for men in Utah who start their adult lives in good health, optimal dynamic contracts close 95-99% of the welfare gap between spot contracts and the "No Borrowing / No Savings Full Insurance" benchmark. However, individuals who are in the poorest health state at age-25 gain much less from dynamic contracts, capturing only 5-29% of this welfare gap.

We also assess the ex ante performance of dynamic contracts by computing the welfare of unborn individuals who face uncertainty about their age-25 health state. While most consumers' benefits from dynamic contracts are similar to those of the healthiest consumers, the low benefits they provide to the sickest consumers dramatically attenuate the benefits of dynamic contracts. That is, the welfare impact of dynamic contracts on unborn individuals is heavily influenced by risk aversion about landing in the worst possible health state at age-25. As a result, for those with flat net income, dynamic contracts close only 43.3% of the welfare gap, while for those with our steepest income profile optimal dynamic contracts bridge only 4.6% of the at-birth "ex ante" welfare gap.

These results suggest that insurance for this pre-age-25 health risk would be desirable in a regime that relied on dynamic health insurance contracts. We find that insuring this pre-age-25 health risk, so that all consumers have the same lifetime welfare as the healthiest 25 year old, would cost the government between \$5,000 and \$12,000 per consumer.[4] We also examine the welfare effect of a balanced budget insurance scheme for pre-age-25 health risk.

[4] One can think of this kind of policy to insure age-25 risk as similar in spirit to the risk-adjustment, risk-corridor, and reinsurance regulations present in the ACA (and many other current environments) but applied to age-25 consumers choosing dynamic contracts instead of consumers of all ages choosing year-to-year contracts.

With this insurance in place, dynamic contracts close between 79-93% of the ex ante welfare gap between spot contracts and our "No Borrowing / No Saving" benchmark.

We also compare the welfare outcomes under full implementation of dynamic contracts to those achieved in an ACA-like exchange. We find that whether the ACA-like exchange is preferred from a welfare perspective depends on the income profiles we consider and whether the government insures pre-age-25 risk under dynamic contracts. With such insurance in place, for rising income profiles, the ACA-like environment does better than dynamic contracts, but for consumers with flat net income profiles, dynamic contracts are preferred. Intuitively, the ACA-like environment is better for individuals who find front-loading costly. If age-25 health risk is not insured under dynamic contracts, then consumers prefer the ACA-like environment for all income paths.

We explore three extensions to our main analysis. First, we consider the effects of switching costs, as estimated in Handel (2013). By enhancing consumers' commitment to their current insurer, these costs improve the insurance of reclassification risk offered in dynamic contracts. This extension can also be viewed as introducing an imperfect credit market, as the switching cost provides a mechanism to insure that consumers repay their debts. Levels of switching costs around the $2,000 estimated in Handel (2013) already lead to meaningful improvements in the welfare performance of dynamic contracts, while a switching cost level of around $10,000 comes close to eliminating reclassification risk entirely.

Second, because consumers who face a spot contracting environment may wish to engage in precautionary savings to reduce the risk they face, we examine the degree to which optimal precautionary savings would improve welfare under spot contracting. While consumer welfare improves, the basic insights from our analysis concerning the benefits of dynamic contracts are unchanged.

Third, because of concerns that estimated levels of risk aversion in the health insurance literature may overstate consumers' aversion to the much larger risks associated with reclassification risk, we also study welfare effects using risk aversion estimates from the macro literature on consumption disasters (Barro (2006)), which are one-fifth the magnitude of the estimate used in our base analysis. Reduced risk aversion limits the extent of surplus that is lost via either spot contracts or dynamic contracts, but, ultimately, the intuition of how dynamic contracting gains relate to income profiles remains the same. Moreover, with this lower level of risk aversion, dynamic contracts close a much larger share (25.6-82.7%) of the at-birth gap between spot contracting and the No-Borrowing/No-Savings benchmark (in the environment with no government insurance of age-25 risk).

In summary, our results show that the value of a system that fully implements optimal dynamic contracts depends crucially on (i) whether there is government insurance of pre-age-25 health risk, and (ii) the steepness of consumer income profiles. A lack of pre-age-25 risk insurance greatly reduces the appeal of dynamic contracts, while, whether or not such insurance is in place, the appeal of these contracts is greater if lifetime income profiles are flatter. With pre-age-25 health risk insurance in place, consumers with flat net income prefer dynamic contracts to the ACA-like environment we study, but consumers with steeper income profiles prefer the ACA-like environment. As well, switching costs and lower risk aversion increase the benefits of dynamic contracts.

In addition to the theoretical work we have noted above, a number of other papers study long-term insurance contracting empirically: Hendel and Lizzeri (2003) examines the structure of life insurance contracts and concludes that they display the features of optimal contracts with one-sided commitment. Finkelstein, McGarry and Sufi (2005) study positive implications of dynamic contracting in the context of long-term care markets, and show evidence of adverse retention, namely that healthier consumers lapse from contracts over time, leading to high average costs from those consumers that remain. Herring and Pauly (2006) conduct an empirical calibration of the Pauly, Kunreuther and Hirth (1995) model, using data from the Medical Expenditure Panel Survey (MEPS). Browne and Hoffmann (2013) study the German private health insurance (PHI) market and demonstrate that (i) front-loading of premiums generates consumer lock-in, (ii) more front-loading is associated with lower lapsation, and (iii) consumers that lapse are healthier than those who do not. Atal (2016) studies the impact of lock-in to an insurance plan on the matching between individuals and health care providers in Chile. Bundorf, Levin, and Mahoney (2012) investigate the implications of reclassification risk in a large-employer context in a short-run environment. Lastly, in a recent working paper, Fleitas, Gowrisankaran, and Losasso (2018) document limited dynamic pass through of expected medical costs into premiums in the pre-ACA small group market. See Hendel (2016) for a survey of the literature on reclassification risk.

Compared to this previous empirical work on long-term health insurance contracts, our paper's contribution is that it (i) uses granular data on a large population to derive optimal contracts with one-sided commitment and (ii) studies the welfare implications of such contracts relative to alternative insurance market designs. In this regard, our paper shares a similar spirit with papers by Farhi and Werning (2013) and Golosov, Troshkin, and Tsyvinski (2016), who characterize and empirically examine the benefits of optimal long-term insurance of individual productivity risk in dynamic macro-style models. Finally, Atal et al (2019) apply our framework to study the long-term health insurance contracts offered in Germany.

## 2 Model of Dynamic Insurance

We consider a dynamic insurance problem $T$ periods long, with periods indexed $t = 1, ...T$. In our empirical analysis, periods represent years, with $t = 1$ corresponding to a 25 year old, and $T = 40$ corresponding to a 65-year old, when Medicare coverage would begin in the U.S.

The consumer may incur medical expenses $m_t \in \mathbb{R}$ in each period $t$, which are uncertain and motivate the desire for insurance. The consumer enters each period $t$ characterized by his health status $\lambda_t \in H$, which determines the distribution of that period's medical expenses. We take $H$ to be a finite set. In our empirical work, greater $\lambda_t$ will indicate sicker individuals, so that expected medical expenses $\mathbb{E}[m_t|\lambda_t]$ are strictly increasing in $\lambda_t$.

The evolution of the consumer's health status is stochastic, with the probability of health status $\lambda_{t+1}$ given previous health history $\Lambda_t^1 \equiv (\lambda_1, ..., \lambda_t)$ given by $\overline{f}(\lambda_{t+1}|\Lambda_t^1)$. (Conditional on $\Lambda_t^1$, the realization of $\lambda_{t+1}$ may be correlated with the realization of $m_t$.) We refer to $\Lambda_t^1$ as the consumer's *health state* at the start of period $t$, and denote the consumer's initial health state by $\Lambda_1^1$. The probabilities $\overline{f}(\cdot)$ give rise to the probability $f(\Lambda_{t'}^{t+1}|\Lambda_t^1)$ that any given continuation path $\Lambda_{t'}^{t+1} \equiv (\lambda_{t+1}, ..., \lambda_{t'})$ of health statuses will follow period $t$ starting from health state $\Lambda_t^1$. In our empirical work, we will suppose that health status transitions are governed by a second-order Markov process, so that $\overline{f}(\cdot)$ can be written as $\overline{f}(\lambda_{t+1}|\lambda_{t-1}, \lambda_t)$, but the results in this section hold more generally.

An individual's health state $\Lambda_t^1$ at the start of period $t$ is observed by both the individual and all insurance firms, namely, there is symmetric information and symmetric learning.[5] We assume that the insurance market is perfectly competitive, with risk-neutral firms who discount future cash flows using the discount factor $\delta \in (0, 1)$. A consumer's risk preferences are described by $u(\cdot)$, the consumer's Bernoulli utility function, while the consumer's long-run expected utility is $\mathbb{E}[\sum_t \delta^t u(c_t)]$, where $c_t \in \mathbb{R}$ is the consumer's period $t$ consumption level. Throughout, we assume that $u'(\cdot) > 0$ and that $u''(\cdot) < 0$, which motivates the consumer's desire for insurance. The consumer's income in period $t$ is $y_t$, and evolves deterministically.[6] Throughout we assume that consumers are unable to borrow to fund premium payments or other expenses.[7]

[5] Our assumption that all insurers have access to the same information assumes that insurers can properly underwrite new customers. If, instead, an individual's current insurer had better information than other firms, prospective insurers would face an adverse selection problem when attempting to attract lapsing consumers. For the consequences of this type of adverse selection, see, for example, DeGaridel-Thoron (2005).

[6] The model readily generalizes to stochastic income, possibly dependent on the consumer's health status. In this case, the optimal contract would insure both health and income risk.

[7] We do, however, introduce switching costs in Sections 2.4 and 8.1, which can be viewed as allowing for an imperfect ability to borrow.

In what follows, we will sometimes refer to a consumer's income profile $y \equiv (y_1, ..., y_T)$ and risk preferences $u(\cdot)$ as the consumer's "type" $\theta \equiv (y, u)$. The optimal contract will depend on this type.

## 2.1 Three Benchmarks

We will compare optimal dynamic contracts with one-sided commitment against three natural benchmarks.[8] The first is the efficient, first-best allocation. In this setting, this outcome involves a constant consumption in all states and periods, equal to the annualized present discounted value of the consumer's "net income" from periods $t = 1$ to $T$ (where the "net income" in period $t$ equals period $t$ income, $y_t$, less the expectation of period $t$ medical expenses conditional on the consumer's health state $\Lambda_1^1$ at the first period of contracting, $\mathbb{E}[m_t|\Lambda_1^1]$). That is, it involves the constant consumption level

$$C^* = \left(\frac{1-\delta}{1-\delta^T}\right) \sum_{t=1}^{T} \delta^{t-1}(y_t - \mathbb{E}[m_t|\Lambda_1^1]). \tag{1}$$

As is well known, if consumers and insurance firms could both commit to a long-term contract given $\Lambda_1^1$, the competitive equilibrium would yield this outcome.

At the opposite extreme, long-term contracts may be impossible, leading to single-period "spot" insurance contracts. In a competitive market, in each period $t$ such contracts will fully insure the consumer's within-period medical expense risks at a premium equal to $\mathbb{E}[m_t|\lambda_t]$, the consumer's expected medical expense given his period $t$ health status $\lambda_t$. This results in the period $t$ consumption level $y_t - \mathbb{E}[m_t|\lambda_t]$. Because the consumer's period $t$ health status $\lambda_t$ is ex ante uncertain, this outcome faces the consumer with risk from an ex ante perspective. Given $\Lambda_1^1$, the consumer's constant certainty equivalent of this uncertain consumption path is the constant consumption level $CE_{SPOT}$ such that

$$u(CE_{SPOT}) = \left(\frac{1-\delta}{1-\delta^T}\right) \mathbb{E}[\sum_{t=1}^{T} \delta^{t-1} u(y_t - \mathbb{E}[m_t|\lambda_t])|\Lambda_1^1] \tag{2}$$

Finally, in this dynamic setting both insurance and consumption smoothing over time are needed to achieve the first best. Since we will focus on settings in which income is increasing over time and borrowing is impossible, another natural benchmark is the outcome that would result if the consumer was fully insured within each period (eliminating all ex ante risk) but resources could not be transferred over time. This certain but time-varying consumption

[8] We also compare the outcome of optimal dynamic contracts with one-sided commitment to an ACA-like exchange market institution in Section 7.

path results in the same welfare as the constant consumption level $C^*_{NBNS}$ ("NBNS" = "No Borrowing/No Savings") such that

$$u(C^*_{NBNS}) = \left(\frac{1-\delta}{1-\delta^T}\right)\sum_{t=1}^{T}\delta^{t-1}u(y_t - \mathbb{E}[m_t|\Lambda_1^1]). \tag{3}$$

Compared to spot contracting, this benchmark eliminates reclassification risk without improving intertemporal allocation.

In our empirical work, we will consider both the expected utility that dynamic contracting and these benchmarks generate at age 25 at the start of contracting (conditional on $\Lambda_1^1$), and also the expected utilities that are implied at birth, factoring in the randomness of the consumer's age-25 health state ($\Lambda_1^1$).[9]

## 2.2 Optimal Dynamic Contracts with One-Sided Commitment

We now turn to the setting in which competitive insurers can offer long-term contracts that they, but not consumers, are committed to. We assume that contracting begins in period 1 (in our empirical setting, at age-25) after $\Lambda_1^1$ has been realized. We can view a long-term contract as specifying the consumer's consumption in each period $t$, $c_t$, as a function of the consumer's publicly-observed health and medical expense history up to through period $t$ including period $t$'s realization of $m_t$ and $\lambda_{t+1}$, $[\Lambda_{t+1}^1, (m_1, ..., m_t)]$.[10] The insurer's profit in the period then equals the consumer's income $y_t$ less the sum of period $t$ medical expenses and period $t$ consumption. The lack of commitment by the consumer, however, means that the consumer is free in each period to change to another insurer who is offering the consumer better terms.

As in Harris and Holmstrom (1982), without loss of generality we can restrict attention when solving for the optimal contract to contracts in which the consumer never has an incentive to "lapse" in this way: since the new contract the consumer signs following any history must give his new insurer a non-negative expected discounted continuation profit, the consumer's initial insurer could include the same contract continuation in the initial insurance contract and weakly increase its expected discounted profit (lapsation would instead yield the

[9] When we examine at-birth welfare levels, we compute $C^*$ and $C^*_{NBNS}$ from an at-birth perspective, replacing $\mathbb{E}[m_t|\Lambda_1^1]$ with $\mathbb{E}[m_t]$ in (1) and (3). For spot contracting, we calculate the certainty equivalent annual consumption that generates the same age 25 to 65 welfare as the spot contracting regime, taking account of any risk aversion losses arising because of uncertainty over the consumer's age-25 health state, $\Lambda_1^1$. Similarly for the dynamic contracts we discuss in the next subsection, and the ACA which we examine in Section 7.

[10] This formulation assumes, for convenience, that the consumer cannot engage in hidden savings. While we will make this assumption initially, in the end we show that under the optimal contract the consumer has no desire to save. We could also allow consumption to be stochastic conditional on $[\Lambda_{t+1}^1, (m_1, ..., m_t)]$, but this will not be optimal.

initial insurer a continuation profit of zero). As a result, we can look for an optimal contract by imposing "lapsation constraints" that require that after no history is it possible to offer the consumer an alternative continuation contract that (i) itself prevents future lapsation, (ii) breaks even in expectation, and (iii) gives the consumer a higher continuation utility than in the original contract.

We take a recursive approach to solving this optimal contracting problem. At each date $t$, we can think of the state as a pair $(\Lambda_t^1, S_t)$ where $\Lambda_t^1$ is the consumer's current health state (which determines future expected medical expenses), and $S_t$ is the absolute value of the loss that the insurer is allowed to sustain going forward (i.e., $S_t$ is the *subsidy* for future insurance).[11] This is a useful formulation for two reasons. First, after any history leading up to period $t$, continuation of the original contract generates some expected utility to the consumer and some expected loss $S_t$ to the insurer. A necessary condition for an optimal contract, given the consumer's current health state, is that it is not possible to increase the consumers' continuation utility while keeping the insurer's loss equal to $S_t$. So, the continuation of the contract must itself solve an optimal contracting problem for an insurer who can sustain the loss $S_t$ starting in health state $\Lambda_t^1$. Second, the constraint that the contract prevents lapsation can be viewed as saying that the consumer's continuation utility starting in any period $t$ when in health state $\Lambda_t^1$ cannot be less than in an optimal contract offered by an insurer who must break even, i.e., who has $S_t = 0$. We denote a contract starting in period $t$ for a consumer in health state $\Lambda_t^1$ by $c_{\Lambda_T^1}(\cdot)$ and the optimal contract for such a consumer whose type is $\theta$ when a subsidy $S_t$ is available by $c_{\Lambda_T^1}^{\theta*}(\cdot|S_t)$.

More formally, consider the problem that arises if a firm faces a consumer of type $\theta$ in health state $\Lambda_t^1$ and can sustain, going forward, a (discounted expected) absolute loss of $S_t$. Let $B^{S_t}(\Lambda_t^1)$ denote the set of period $t$ contracts in health state $\Lambda_t^1$ that break even in expectation with a subsidy of $S_t$ if no lapsation occurs, and let $V_{\Lambda_t^1}^{\theta}(c_{\Lambda_t^1}(\cdot))$ be the consumer's discounted expected utility from contract $c_{\Lambda_t^1}(\cdot)$ starting in period $t$ with health state $\Lambda_t^1$. In addition, let $c_{\Lambda_t^1|\Lambda_{t'}^{t+1}}(\cdot)$ denote the continuation of contract $c_{\Lambda_t^1}(\cdot)$ starting in period $t' > t$ if the health status realizations between period $t+1$ and $t'$ are $\Lambda_{t'}^{t+1} \equiv (\lambda_{t+1}, ..., \lambda_{t'})$, resulting in a period $t$ health state of $\langle \Lambda_t^1, \Lambda_{t'}^{t+1} \rangle \equiv (\lambda_1, ..., \lambda_t, \lambda_{t+1}, ..., \lambda_{t'})$. The optimal contract for a consumer of type $\theta$ is then described as follows:

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

$$\max_{c_{\Lambda_t^1}(\cdot) \in B^{S_t}(\Lambda_t^1)} V_{\Lambda_t^1}^{\theta}(c_{\Lambda_t^1}(\cdot)) \tag{4}$$

[11] Note that because only the health state $\Lambda_t^1$ matters for the distribution of future medical expenses, an optimal contract will not depend on previous medical expense realizations.

$$s.t.\;\; V^{\theta}_{\langle \Lambda^1_t, \Lambda^{t+1}_{t'} \rangle}(c_{\Lambda^1_t|\Lambda^{t+1}_{t'}}(\cdot)) \geq V^{\theta}_{\langle \Lambda^1_t, \Lambda^{t+1}_{t'} \rangle}(c^{\theta *}_{\langle \Lambda^1_t, \Lambda^{t+1}_{t'} \rangle}(\cdot|0)) \; \textit{for all } \Lambda^{t+1}_{t'} \textit{ with } t' > t$$

Note that problem (4) 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 $c^{\theta *}_{\langle \Lambda^1_t, \Lambda^{t+1}_{t'} \rangle}(\cdot|0)$, the optimal contract starting at health state $\langle \Lambda^1_t, \Lambda^{t+1}_{t'} \rangle$ with no subsidy. The constraint ensures us that, following the realization of continuation health history $\Lambda^{t+1}_{t'}$, 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 and that also satisfies no-lapsation.

Our main result, which can be viewed as a generalization of Harris and Holmstrom (1982), and which we establish in the Appendix, is:[12]

**Proposition 1** *The optimal contract for a consumer of type $\theta$ starting in period $t$ at health state $\Lambda^1_t$, denoted by $c^{\theta *}_{\Lambda^1_1}(\cdot)$, is fully characterized by the zero-profit condition and, for all $t' > t$ and $\Lambda^{t+1}_{t'}$ such that $f(\Lambda^{t+1}_{t'}|\Lambda^1_t) > 0$, the condition that the consumer receives the following certain consumption level:*

$$c^{\theta *}_{\Lambda^1_t}(\langle \Lambda^1_t, \Lambda^{t+1}_{t'} \rangle) = \max\{c^{\theta *}_{\Lambda^1_t}(\Lambda^1_t), \max_{\tau \in \{t+1,\ldots,t'\}} c^{\theta *}_{\langle \Lambda^1_t, \Lambda^{t+1}_{\tau} \rangle}(\langle \Lambda^1_t, \Lambda^{t+1}_{\tau} \rangle)\}. \tag{5}$$

*Under this contract, the consumer does not wish to secretly save.*

In words, the optimal contract $c^{\theta *}_{\Lambda^1_t}(\cdot)$ signed in period $t$ at health state $\Lambda^1_t$ offers in each period $t' > t$ after 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 $\Lambda^{t+1}_{t'}$. Thus, applying this result to the initial contracting in period 1, the optimal contract starting in period 1 offers an initial *consumption floor*, which is then bumped up in later periods $t > 1$ each time the consumer reaches a state in which the market would offer a higher initial consumption floor. The equilibrium contract provides full within-period insurance for the consumer (i.e., consumption in each period is independent of $m_t$), and partial insurance against reclassification risk, as consumers who have experienced sufficiently bad health states leading up through a given period $t$ [i.e., such that $c^{\theta *}_{\Lambda^1_\tau}(\Lambda^1_\tau) \leq c^{\theta *}_{\Lambda^1_1}(\Lambda^1_1)$ for all $\tau \leq t$] all enjoy the same level of consumption regardless of differences in their period $t$ health states. The extent of this partial insurance is a function of the consumer's initial health state

[12] Aside from the change of context from labor markets to health insurance, our model generalizes Harris and Holmstrom (1982) by allowing for more general stochastic medical expense and health state transition processes. We also offer a recursive proof of the result, show that the optimal contracts are self-selective (Section 2.3), and introduce switching costs (Sections 2.4 and 8.1).

$\Lambda_1^1$, income profile $y$, and risk preferences $u(\cdot)$, as well as the health state transition process. Since the consumer's consumption level is always weakly rising over time, the consumer never wishes to save.[13]

The optimal contract in Proposition 1 specifies the consumer's consumption levels for each possible health history and prevents lapsation. Note, however, that the same outcome can alternatively be achieved by means of a much simpler *guaranteed premium path contract* from which the consumer may lapse. Specifically, the consumer is given the option to renew, if he has not yet lapsed, at the guaranteed premium path $p^{\theta*}(\Lambda_1^1) \equiv (p_1^{\theta*}, ..., p_T^{\theta*})$ where $p_t^{\theta*} = y_t - c_{\Lambda_1^1}^{\theta*}(\Lambda_1^1)$ for $t = 1, ...T$, provided that he has always renewed in the past.[14] That is, the guaranteed premium path keeps consumption constant over time, equal to $c_{\Lambda_1^1}^{\theta*}(\Lambda_1^1)$, as long as the consumer sticks with the contract. But if the consumer arrives in a period $t$ with a sufficiently good health state $\Lambda_t^1$, he may choose not to renew, instead signing a contract with a new insurer that offers guaranteed premium path $p^{\theta*}(\Lambda_t^1) \equiv \{y_\tau - c_{\Lambda_t^1}^{\theta*}(\Lambda_t^1)\}_{\tau \geq t}$ where $c_{\Lambda_t^1}^{\theta*}(\Lambda_t^1) > c_{\Lambda_1^1}^{\theta*}(\Lambda_1^1)$. Such lapses have no effect on the profit of the consumer's initial insurer as that firm was indifferent about whether to match the outside offer.[15,16]

## 2.3 Unobserved Types and Self-Selection

The analysis above assumed that a consumer's lifetime income profile $y = (y_1, ..., y_T)$ and risk preferences $u(\cdot)$, summarized in the consumer's "type" $\theta = (y, u)$, were known by both the consumer and all insurers. In reality, this is unlikely to be the case, which could, in principle, pose an important obstacle to these contracts' practical use. In this subsection we show that insurers' failure to possess this information poses no such problem. Specifically, we show that if offered the collection of optimal contracts for all types derived above, presented as

[13] For a formal derivation, which requires specifying what happens if a consumer with hidden savings seeks to buy insurance from a new firm, see Appendix B.

[14] This form of contract is the counterpart to the "Annual Renewable Term" life insurance contracts studied in Hendel and Lizzeri (2003).

[15] We discuss in Section 8 the difference between this guaranteed premium path contract and that of Pauly, Kunreuther, and Hirth (1995).

[16] The recursive formulation also makes clear that this equilibrium outcome can be achieved instead with single-period contracts. A consumer in period 1 with health state $\Lambda_1^1$ could purchase a contract that covers all period 1 medical expenses, and that in addition pays the consumer at the start of period 2 the amount that the optimal contract implicitly subsidizes the realized continuation state $\Lambda_2^1$. This amount would allow the consumer to buy the long-term continuation contract on the open market. Upon reaching period 2, however, the consumer could instead again buy a one-period policy of this type, and could continue in this manner until period $T$. [This approach to replicating a long-term contract with a series of short-term contracts is reminiscent of Fudenberg et al. (1990), although our setting is not captured in their model because of the presence of lapsation constraints and the consumer's inability to borrow.]

As noted in Cochrane (1995), such short-term contracts avoid the consumer being locked into an insurer, perhaps resulting in better insurer performance as well as better matching of insurers and consumers when health care networks are bundled with insurance provision. However, such contracts require that courts can verify the consumer's health state $\Lambda_t^1$, while guaranteed premium path contracts do not.

guaranteed premium path contracts, consumers will self-select, choosing the optimal contract for their type.[17]

Specifically, suppose that there is a set $\Theta$ of types in the market.[18] As above, a guaranteed premium path contract is a $p = (p_1, ..., p_T)$ that allows the consumer to continue coverage in period $t$ paying premium $p_t$ provided that he has not previously lapsed. As described above, the optimal guaranteed premium path contract for a known type $\theta$ starting in period $t$ when the consumer's health state is $\Lambda_t^1$ is denoted by the path $p^{\theta *}(\Lambda_t^1) \equiv \{y_\tau - c^{\theta *}_{\Lambda_t^1}(\Lambda_t^1)\}_{\tau \geq t}$, a path that keeps consumption constant [equal to $c^{\theta *}_{\Lambda_t^1}(\Lambda_t^1)$] as income changes from year to year.

Our result is:

**Proposition 2** *Suppose that, in each period $t = 1, ..., T$, the menu of optimal guaranteed premium path contracts $\{p^{\theta *}(\Lambda_t^1)\}_{\theta \in \Theta}$ is offered to a consumer in health state $\Lambda_t^1$, where $p^{\theta *}(\Lambda_t^1) \equiv \{y_\tau - c^{\theta *}_{\Lambda_t^1}(\Lambda_t^1)\}_{\tau = t}^T$. Then in each period the menu is self-selective and induces no secret savings: that is, if a consumer of type $\theta$ agrees to a new contract he chooses that type's optimal contract $p^{\theta *}(\Lambda_t^1)$ and does not secretly save.*

**Proof.** In Appendix B. ■

Since insurers cannot offer any type of consumer a greater value than in the optimal contract and still break even, Proposition 2 implies that it is an equilibrium for this menu of contracts to be offered, which results in the same allocation as if consumer types were perfectly observable.

## 2.4 Switching Costs

Recent evidence suggests that consumers may exhibit substantial inertia in their health insurance choices [see, e.g., Handel (2013)]. We can extend our analysis to consider the effect of switching costs, modeled as creating a consumption loss of $\sigma > 0$ if the consumer lapses and switches insurers. The key change this introduces is that the inequality in the lapsation constraint in the period $t$ problem (4) becomes $V^\theta_{\langle \Lambda_t^1, \Lambda_{t'}^{t+1} \rangle}(c_{\Lambda_t^1 | \Lambda_{t'}^{t+1}}(\cdot)) \geq V^\theta_{\langle \Lambda_t^1, \Lambda_{t'}^{t+1} \rangle}(c^{\theta *}_{\langle \Lambda_t^1, \Lambda_{t'}^{t+1} \rangle}(\cdot | - \sigma))$; an insurer seeking to induce the consumer to lapse must now incur the cost $\sigma$ to compensate the consumer for his switching cost. Nonetheless, as we show in Appendix A, the basic structure of an optimal contract is unchanged when switching

[17] Note that contracts that instead present the optimal contracts as guaranteed consumption levels (as in Proposition 1), would clearly not induce self-selection as consumers with low lifetime incomes would choose contracts intended for consumers with high lifetime incomes.

[18] Formally, we allow $\Theta$ to include all possible income paths $y \in \mathbb{R}_+^T$ to allow for the possibility of secret savings (see footnote 61 in Appendix B).

costs are present, again involving consumption guarantees. Switching costs simply allow those guarantees to be greater because healthy consumers are less likely to need to receive a premium reduction (consumption increase) to prevent lapsation, enabling a greater shift of resources from healthy to unhealthy states. They also allow for better consumption smoothing over time, as less front-loading is needed to prevent lapsation.[19]

# 3 Data and Parameter Estimates

We investigate positive and normative outcomes for each type of contracting situation. To predict equilibrium contracts and welfare under each regime we need four basic ingredients: (i) expected medical costs conditional on an individual's health status, (ii) the transitions across health states as individuals age, (iii) preferences towards risk, and (iv) income profiles.

We focus on the sample of men that appear in the all-payer claims data from the State of Utah for the years 2013-2015.[20] This dataset includes detailed medical claims for each individual in the state of Utah except for those individuals enrolled in traditional Medicare or those who are uninsured. Our analysis studies all men in the data who (i) are 25-64 years old throughout the three-year sample period and (ii) appear in the data each month throughout the sample period (e.g., they have no spells of non-insurance). We make the latter restriction in order to ensure that we can cleanly capture health status transitions, as described in more detail later.

Table 1 describes our final sample of Utah men, with key descriptive statistics broken down by age. Our sample has 212,265 men who averaged $4,650 in total medical spending in 2015. Not surprisingly, average total medical spending is increasing with age.

## 3.1 Health States

The most essential part of the data is the available information on the diagnostics (ICD-9 codes) of each individual in the sample. We feed the diagnostic codes as well as other demographics into the ACG software developed at Johns Hopkins Medical School to create individual-level measures of predicted expected medical expenses for the upcoming year relative to the mean of the population.[21] The output is an index that represents the health

[19] Indeed, these switching costs can be viewed as allowing for a limited amount of borrowing by the consumer.

[20] These data are utilized as well in Lavetti et al. (2018), which contains a more complete description of the data.

[21] We use the Johns Hopkins ACG (Adjusted Clinical Groups) Case-Mix System. It is one of the most widely used and respected risk adjustment and predictive modeling packages in the health care sector, specifically designed to use diagnostic claims data to predict future medical expenditures.

| Sample and Total Health Expenditure | | |
|---|---|---|
| Ages | Population | Mean |
| All | 212,265 | 4,650 |
| 25-29 | 14,872 | 2,302 |
| 30-34 | 26,412 | 2,520 |
| 35-39 | 31,414 | 2,871 |
| 40-44 | 29,864 | 3,424 |
| 45-49 | 27,424 | 4,259 |
| 50-54 | 31,082 | 5,659 |
| 55-59 | 31,574 | 7,339 |
| 60-64 | 17,131 | 8,964 |

Table 1: Sample statistics for (i) the entire sample of men aged 25-64 used in our equilibrium analysis and (ii) 5-year age buckets within that sample. For each relevant group, "Population" column reports the number of individuals, and "Mean" column reports the average medical cost in 2015.

status of each individual in the population. Since the ACG is used by insurers in their underwriting processes, our empirics are based on similar information about risks that market participants (insurers) have.[22] We denote the ACG index by $\lambda$ and we refer to $\lambda_{it}$ as individual $i$'s "health status" at time $t$.

To ensure meaningful support when we estimate transition matrices, we partition the health statuses into seven mutually exclusive and exhaustive bins that each contain one-seventh of the final sample. Table 2 shows the proportion of individuals in each age group in each of these seven health categories, with bin 1 being the healthiest, and bin 7 being the sickest. In its last row it also shows the expected expenses corresponding to each bin: an individual in the healthiest bin has expected annual medical expenses of \$837, while someone in the sickest bin has expected annual medical expenses of \$20,507.

## 3.2 Health State Transitions

The second key input into our empirical analysis is health transitions over time. We model transitions in health status as a second-order Markov process in which the distribution of an individual $i$'s period $t+1$ status $\lambda_{i,t+1}$ is conditional on his health status in the previous two years, $\Lambda_t \equiv (\lambda_{i,t-1}, \lambda_{it})$.[23] Specifically, once we have $\lambda_{it}$ for every individual and year in the sample, we estimate year-to-year transition probabilities $\overline{f}(\lambda_{i,t+1}|\lambda_{i,t-1}, \lambda_{it})$ for individuals in five-year age groups (e.g., transitions within cohort 25-30) using the actual transitions

[22] This is one of the main advantages of our empirical strategy. Most of the literature on health insurance estimates the distribution of risks from observed insurance choices and realized total medical expenditures. Instead our measure of risk is based on diagnosis codes and professional diagnostics (the ACG index).

[23] We are limited by our data to modeling the dependency of transitions based only on the last two years.

| Health Status by Age | | | | | | | |
|---|---|---|---|---|---|---|---|
| Age | 1 (Healthy) | 2 | 3 | 4 | 5 | 6 | 7 (Sick) |
| 25-29 | 0.55 | 0.15 | 0.11 | 0.07 | 0.05 | 0.04 | 0.05 |
| 30-34 | 0.44 | 0.17 | 0.14 | 0.08 | 0.06 | 0.06 | 0.05 |
| 35-39 | 0.37 | 0.17 | 0.14 | 0.10 | 0.08 | 0.07 | 0.07 |
| 40-44 | 0.31 | 0.16 | 0.15 | 0.12 | 0.10 | 0.08 | 0.08 |
| 45-49 | 0.17 | 0.12 | 0.18 | 0.16 | 0.14 | 0.12 | 0.11 |
| 50-54 | 0.11 | 0.10 | 0.15 | 0.17 | 0.17 | 0.15 | 0.15 |
| 55-59 | 0 | 0.11 | 0.10 | 0.18 | 0.19 | 0.21 | 0.20 |
| 60-64 | 0 | 0.08 | 0.08 | 0.17 | 0.19 | 0.23 | 0.25 |
| Actuarial costs | 837 | 1,376 | 1,973 | 3,052 | 4,356 | 6,840 | 20,507 |

Table 2: Health status by age in 2015 for our sample, where consumers are divided into 7 bins of their predicted medical spending (determined by their Johns Hopkins ACG predictive score) for the year ahead. The "Actuarial costs" row reports how expected expenses in the upcoming year vary across the consumer health status bins.

of consumers within each age range. (Again, the five-year grouping helps ensure adequate populations in each cell.) The advantage of computing transitions of ACG scores as opposed to medical expense transitions is that the ACG is based on persistent diagnostics. A broken arm probably does not affect significantly future medical expense realizations while asthma does. In other words, the ACG eliminates temporary expenses from the forecast of future expenses.[24] From the estimated probabilities $f(\lambda_{i,t+1}|\lambda_{i,t-1}, \lambda_{it})$ we construct the 49-by-49 health state transition matrices [giving the mapping from $(\lambda_{i,t-1}, \lambda_{i,t})$ to $(\lambda_{i,t}, \lambda_{i,t+1})$] for the five-year age bins from ages 25-65 as the foundation for modeling health state persistence and transitions over time.

Tables 3 and 4 present the estimates of $f(\lambda_{i,t+1}|\lambda_{i,t-1}, \lambda_{it})$ for ages 30-35 and 40-45 respectively (rounded to the nearest 0.01). Panel A shows transitions from states in which an individual's health statuses were the same in each of the two previous years (i.e., states in which $\lambda_{i,t-1} = \lambda_{it}$). Entries along the diagonal of each matrix reflect health state persistence, while off-diagonal elements reflect health state changes. For example, 73% of consumers aged 30-35 who are in the healthiest possible state ($\lambda_{i,t+1} = \lambda_{it} = 1$) are estimated to stay in the healthiest health status for the following year. Only 6% of these consumers begin the following year in one of the four sickest bins ($\lambda_{i,t+1} \in \{4, 5, 6, 7\}$). On the other hand, 89% of consumers who are in the worst possible health state ($\lambda_{i,t-1} = \lambda_{it} = 7$) begin the following year in one of the two sickest bins ($\lambda_{i,t+1} \in \{6, 7\}$). The next two panels (B and C) show the

[24] Admittedly, by defining transitions over ACGs we may miss potential information on what condition led to the current ACG index that could entail different persistence beyond two years. However, we believe that with the combination of using ACG scores rather than medical expenses, and two-year health states, we capture the health transition process reasonably well.

probabilities $\overline{f}(\lambda_{i,t+1}|\lambda_{i,t-1},\lambda_{it})$ when the consumer has had the worst possible health status in at least one of the two previous years. Comparing the case where $(\lambda_{i,t-1},\lambda_{it}) = (1,7)$ to the case in which $(\lambda_{i,t-1},\lambda_{it}) = (7,1)$, we see that the previous year's health status has greater importance than health status two years prior. Though the distributions of health are different for 40-45 year olds, health states show similar persistence. [25]

The persistence embodied in these health state transitions is illustrated in Tables 5 and 6. Table 5 shows the net present value of expected medical expenses for different future periods conditional on the consumer's age 30 health state, focusing on health states in which the consumer's health status was the same at ages 29 and 30. Table 6 shows the same kind of calculation for consumers starting at age 45. The tables show that while there is significant persistence, much (but not all) of it dissipates after 10 years.[26,27]

## 3.3 Risk Preferences

The third ingredient is a consumer's risk aversion, i.e., the degree to which consumption smoothing over different states of the world is valued by consumers. In our main analysis, we use the risk preferences estimated in Handel, Hendel, and Whinston (2015). There we estimate a panel discrete choice model where risk aversion is identified by the choices that households make conditional on their household-specific health expenditure risk for the upcoming year. Consumers have constant absolute risk aversion (CARA) preferences:

$$u(c) = -\frac{1}{\gamma}e^{-\gamma c} \tag{6}$$

where $c = y - p - o$ is consumption (which equals income $y$ less premium payments $p$ and out-of-pocket medical expenses $o$) and $\gamma$ is the risk aversion parameter. The mean estimated risk-aversion level is 0.0004, which falls within the range reported in the literature. We also consider the robustness of our conclusions with respect to the degree of risk aversion in Section 8.3.

[25] Lavetti et al. (2018) presents transition matrix estimates that are similar in spirit to those used here, and has the matrices we use here available upon request.

[26] The fact that expected costs depend relatively little on the health state 10 years prior is consistent with actuarial mortality tables. There are two kinds of tables: "ultimate" tables are based on attained age only, while "select and ultimate" tables report the death rate not only by attained age, but by years since underwriting (namely, conditional on being in good health at that time). The tables converge as the years since underwriting increase; 10 years after underwriting the rates are quite similar.

[27] Note that complete persistence would eliminate the benefit of dynamic contracts as there would be no reclassification risk to insure once a consumer's age-25 health state is realized.

| | | $\lambda_{t+1}$ | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $\lambda_{t-1}$ | $\lambda_t$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | 1 | 0.73 | 0.13 | 0.07 | 0.02 | 0.02 | 0.01 | 0.01 |
| 2 | 2 | 0.41 | 0.27 | 0.16 | 0.07 | 0.04 | 0.03 | 0.02 |
| 3 | 3 | 0.20 | 0.21 | 0.28 | 0.17 | 0.08 | 0.05 | 0.02 |
| 4 | 4 | 0.08 | 0.09 | 0.19 | 0.32 | 0.15 | 0.11 | 0.05 |
| 5 | 5 | 0.07 | 0.06 | 0.09 | 0.15 | 0.37 | 0.21 | 0.06 |
| 6 | 6 | 0.04 | 0 | 0.04 | 0.03 | 0.15 | 0.51 | 0.23 |
| 7 | 7 | 0.02 | 0 | 0.02 | 0.02 | 0.05 | 0.12 | 0.77 |

Panel A: Health-state transition probabilities for $\lambda_{t-1} = \lambda_t$ health states

| | | $\lambda_{t+1}$ | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $\lambda_{t-1}$ | $\lambda_t$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | 1 | 0.73 | 0.13 | 0.07 | 0.02 | 0.02 | 0.01 | 0.01 |
| 1 | 2 | 0.52 | 0.23 | 0.12 | 0.05 | 0.04 | 0.03 | 0.02 |
| 1 | 3 | 0.42 | 0.21 | 0.18 | 0.07 | 0.04 | 0.04 | 0.02 |
| 1 | 4 | 0.34 | 0.19 | 0.15 | 0.13 | 0.09 | 0.07 | 0.03 |
| 1 | 5 | 0.27 | 0.22 | 0.15 | 0.11 | 0.09 | 0.09 | 0.06 |
| 1 | 6 | 0.39 | 0.18 | 0.15 | 0.07 | 0.07 | 0.09 | 0.05 |
| 1 | 7 | 0.30 | 0.18 | 0.19 | 0.06 | 0.06 | 0.09 | 0.12 |

Panel B: Health-state transition probabilities for $\lambda_{t-1} = 1$ health states

| | | $\lambda_{t+1}$ | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $\lambda_{t-1}$ | $\lambda_t$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 7 | 1 | 0.63 | 0.25 | 0.12 | 0 | 0 | 0 | 0 |
| 7 | 2 | 0.52 | 0.26 | 0.22 | 0 | 0 | 0 | 0 |
| 7 | 3 | 0.16 | 0.18 | 0.29 | 0.14 | 0.13 | 0.10 | 0 |
| 7 | 4 | 0.15 | 0.16 | 0.21 | 0.20 | 0.14 | 0.13 | 0 |
| 7 | 5 | 0 | 0 | 0.11 | 0.22 | 0.26 | 0.30 | 0.12 |
| 7 | 6 | 0 | 0 | 0.08 | 0.08 | 0.13 | 0.37 | 0.34 |
| 7 | 7 | 0.02 | 0 | 0.02 | 0.02 | 0.05 | 0.12 | 0.77 |

Panel C: Health-state transition probabilities for $\lambda_{t-1} = 7$ health states

Table 3: Health status transitions from one year to the next, for 30-35 year old men.

| $\lambda_{t-1}$ | $\lambda_t$ | $\lambda_{t+1}$ 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0.66 | 0.15 | 0.09 | 0.04 | 0.02 | 0.02 | 0.01 |
| 2 | 2 | 0.29 | 0.29 | 0.20 | 0.11 | 0.05 | 0.03 | 0.03 |
| 3 | 3 | 0.12 | 0.16 | 0.32 | 0.20 | 0.10 | 0.05 | 0.03 |
| 4 | 4 | 0.05 | 0.08 | 0.15 | 0.35 | 0.20 | 0.11 | 0.07 |
| 5 | 5 | 0.03 | 0.05 | 0.07 | 0.17 | 0.35 | 0.23 | 0.10 |
| 6 | 6 | 0.03 | 0.03 | 0.03 | 0.05 | 0.14 | 0.49 | 0.23 |
| 7 | 7 | 0.01 | 0.01 | 0.02 | 0.04 | 0.05 | 0.13 | 0.74 |

Panel A: Health-state transition probabilities for $\lambda_{t-1} = \lambda_t$ health states

| $\lambda_{t-1}$ | $\lambda_t$ | $\lambda_{t+1}$ 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0.66 | 0.15 | 0.09 | 0.04 | 0.02 | 0.02 | 0.01 |
| 1 | 2 | 0.40 | 0.25 | 0.18 | 0.08 | 0.05 | 0.03 | 0.02 |
| 1 | 3 | 0.30 | 0.23 | 0.23 | 0.12 | 0.07 | 0.03 | 0.03 |
| 1 | 4 | 0.22 | 0.18 | 0.22 | 0.18 | 0.09 | 0.07 | 0.05 |
| 1 | 5 | 0.23 | 0.16 | 0.17 | 0.17 | 0.13 | 0.09 | 0.04 |
| 1 | 6 | 0.28 | 0.16 | 0.17 | 0.11 | 0.13 | 0.10 | 0.06 |
| 1 | 7 | 0.28 | 0.13 | 0.16 | 0.11 | 0.09 | 0.10 | 0.13 |

Panel B: Health-state transition probabilities for $\lambda_{t-1} = 1$ health states

| $\lambda_{t-1}$ | $\lambda_t$ | $\lambda_{t+1}$ 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| 7 | 1 | 0.59 | 0.18 | 0.23 | 0 | 0 | 0 | 0 |
| 7 | 2 | 0.30 | 0.34 | 0.36 | 0 | 0 | 0 | 0 |
| 7 | 3 | 0.12 | 0.13 | 0.31 | 0.26 | 0.11 | 0 | 0.07 |
| 7 | 4 | 0 | 0.10 | 0.19 | 0.24 | 0.22 | 0.13 | 0.12 |
| 7 | 5 | 0 | 0 | 0.11 | 0.23 | 0.27 | 0.25 | 0.14 |
| 7 | 6 | 0 | 0 | 0 | 0.07 | 0.18 | 0.41 | 0.33 |
| 7 | 7 | 0.01 | 0.01 | 0.02 | 0.04 | 0.05 | 0.13 | 0.74 |

Panel C: Health-state transition probabilities for $\lambda_{t-1} = 7$ health states

Table 4: Health status transitions from one year to the next, for 40-45 year old men.

| Health status | Ages | | | |
|---|---|---|---|---|
| (age 29 and 30) | 30 | 31-35 | 36-40 | 41-64 |
| 1 | 0.84 | 1.83 | 2.39 | 4.2 |
| 2 | 1.37 | 2.22 | 2.58 | 4.24 |
| 3 | 1.97 | 2.83 | 2.87 | 4.28 |
| 4 | 3.05 | 3.98 | 3.41 | 4.37 |
| 5 | 4.36 | 4.77 | 3.81 | 4.43 |
| 6 | 6.84 | 8.35 | 5.26 | 4.65 |
| 7 | 20.51 | 13.05 | 6.31 | 4.78 |

Table 5: This table 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.

| Health status | Ages | | | |
|---|---|---|---|---|
| (age 44 and 45) | 45 | 46-50 | 51-55 | 56-64 |
| 1 | 0.84 | 2.27 | 4.07 | 5.35 |
| 2 | 1.37 | 3.17 | 4.54 | 5.51 |
| 3 | 1.97 | 3.62 | 4.77 | 5.58 |
| 4 | 3.05 | 4.51 | 5.19 | 5.70 |
| 5 | 4.36 | 5.74 | 5.73 | 5.85 |
| 6 | 6.84 | 8.22 | 6.59 | 6.07 |
| 7 | 20.51 | 13.05 | 7.90 | 6.38 |

Table 6: This table 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-45 health states.

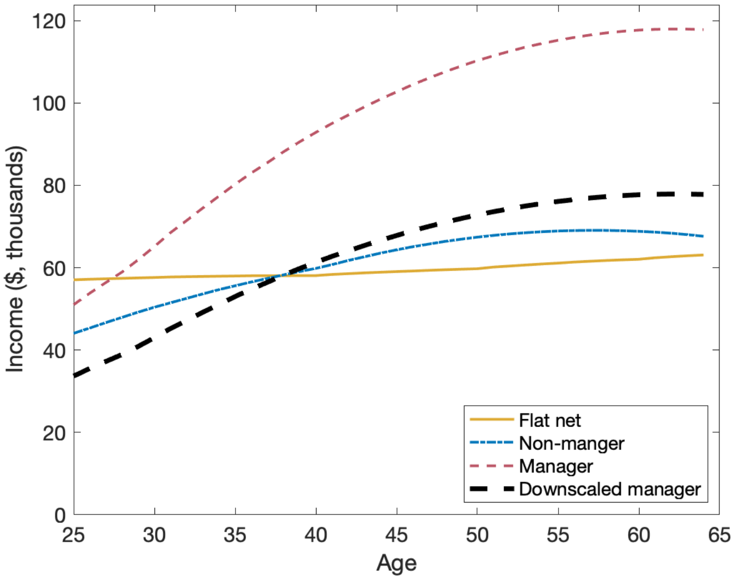
### 3.4 Income profiles

The shape of the optimal contract depends on a consumer's income profile. Insurers offer different contracts to consumers with different income profiles to maximize their lifetime expected utilities conditional on breaking even and the lapsation constraint. We study this effect by computing optimal contracts and welfare for several different income profiles that vary in how steeply income rises with age. The least steep is a *flat net income profile*, in which the change in income each year equals the change in the population's average medical expenses. With this income profile, there are no intertemporal consumption smoothing motives (for an individual who would pay a premium in each period equal to the population average medical costs), as individuals with flat net income do not want to use the contract as a mechanism to borrow or save, unlike consumers with increasing or decreasing income profiles over time.[28] We also examine several more steeply rising income profiles, based on the income profiles we observed in Handel, Hendel, and Whinston (2015) for managers and non-managers in the firm studied there. Income profiles of managers at the firm were steepest, while those of non-managers were flatter but still steeper than a flat net income profile. Figure 3.4 shows the manager income profile as the highest dashed curve, while non-managers' income profile is represented by the flatter curve comprised of long and short dashes. The bold dashed curve in the figure is a proportionally scaled-down managers' income profile that makes the present value of lifetime income equal to that of a non-manager (which facilitates certain comparative statics we present). The solid curve is a flat net income profile with the same net present value as the non-manager and downscaled-manager profiles.

# 4 Computation of Optimal Dynamic Contracts

Given estimated values of transitions $\overline{f}(\cdot)$ and preferences $u(\cdot)$, combined with an income profile $y$ and a value for the discount factor $\delta$ which we set to 0.975, we derive recursively the contracts described in Proposition 1, which characterized the contracts that would be offered to consumers in a competitive market with one-sided commitment. For ease of notation, in this section we denote the initial consumption guarantee offered to a consumer signing a contract in period $t$ with health state $\Lambda_t = (\lambda_{t-1,}\lambda_t)$ by $c_t^{\theta *}(\Lambda_t)$. For a consumer of type $\theta$, we first derive, for each possible last period state $\Lambda_T = (\lambda_{T-1,}\lambda_T)$, the last period consumption levels $c_T^{\theta *}(\Lambda_T) = y_T - \mathbb{E}[m_T|\lambda_T]$ that would be offered to a consumer in state $\Lambda_T$ by competitive firms. We then look at each possible state $\Lambda_{T-1} = (\lambda_{T-2,}\lambda_{T-1})$ in period $T-1$. We find $c_{T-1}^{\theta *}(\Lambda_{T-1})$ by doing a binary search over possible values for the consumption guarantee

[28] The population average medical costs over time are close to the expected medical costs over time of a consumer conditional on being in the healthiest state at age 25. Thus, such a consumer has little incentive to borrow or save if they have a flat net income profile.




$c_{T-1}$, looking for the largest $c_{T-1}$ that generates non-negative profits for the insurer, taking account of the fact that the consumer will lapse in period $T$ (or, if retained, yield the insurer continuation profits of zero) in those states $\Lambda_T$ in which $c_T^{\theta*}(\Lambda_T) > c_{T-1}$. We then continue backward in this fashion, with the transitions $f(\cdot)$ being used to generate probabilities that the consumer is in each possible state at each future date (which also generates the probability that the consumer will have lapsed by that date).

More formally, enumerate the 49 possible combinations of $(\lambda_{t-1,}\lambda_t)$ for each period $t$ by $\{\Delta^s = (\lambda_{t-1,}^s\lambda_t^s)\}_{s=1}^{49}$. For each period $t$, we denote by $C_t^\theta$ the $(T-t+1)\times 49$ matrix of first-period consumption guarantees whose $(\tau, s)$ element for $\tau \geq t$ and $s \in \{1, ..., 49\}$ is $c_\tau^{\theta*}(\Lambda^s)$, a consumption guarantee that breaks even for a contract starting in period $\tau$ with health state $\Lambda_\tau = \Lambda^s$, given the future guarantees described in Proposition 1 (which are, themselves, contained in $C_{t+1}^\theta$). We start at $t = T$ where, as noted above, $c_T^{\theta*}(\Lambda_T = \Lambda^s) = y_T - \mathbb{E}[m_T|\lambda_t^s]$. This gives us $C_T^\theta$. We then proceed iteratively backwards, deriving $C_t^\theta$ given $C_{t+1}^\theta$ and the transition probabilities. Specifically, $C_t^\theta$ adds an additional row to $C_{t+1}^\theta$; each element $(t, s)$ of this row is the consumption guarantee $c_t^{\theta*}(\Lambda_t = \Lambda^s)$, which we derive by doing a grid search to find the (unique) value $\bar{c}$ such that

$$\{y_t - \mathbb{E}(m_t|\lambda_t = \lambda_t^s) - \bar{c}\} + \left\{ \sum_{\tau=t+1}^{T} \delta^{\tau-t} \sum_{z=1}^{49} [y_\tau - \mathbb{E}(m_\tau|\lambda_\tau = \lambda_t^z) - \bar{c}] \cdot P_\tau(z|\Lambda_t = \Lambda^s, C_{t+1}^\theta, \bar{c}) \right\} = 0, \tag{7}$$

where $P_\tau(z|\Lambda_t = \Lambda^s, C_{t+1}^\theta, \bar{c})$ is the probability that, starting in health state $\Lambda_t = \Lambda^s$ in period $t$, the health state transitions $(\Lambda_{t+1}, ..., \Lambda_\tau)$ from period $t$ to period $\tau$ are such that

$\Lambda_\tau = \Lambda^z$ and $c_{t'}^{\theta*}(\Lambda_{t'}) \leq \overline{c}$ for all $t' \in \{t+1, ..., \tau\}$.[29] The first term in curly brackets in (7) is the expected profit in period $t$ given the initial consumption guarantee $\overline{c}$, while the second term in curly brackets is the expected continuation profit and uses the fact that starting at the first instance at which the consumer's health state $\Lambda_\tau$ is such that $c_\tau^{\theta*}(\Lambda_\tau) > \overline{c}$ (i.e., the lapsation constraint binds), the insurer's continuation payoff is zero. We continue in this iterative manner until we have derived $C_1^\theta$, whose first row gives the initial consumption guarantees offered in period 1 to consumers in each of the 49 possible period-1 health states.

# 5 Results: Optimal Contracts

Using the data and computational approach described above, we find the consumptions and premiums in optimal dynamic contracts with one-sided commitment, and then compare their outcome to those in various benchmarks. In this section, we first study the structure of these contracts, examining the extent to which they are front-loaded, the degree of reclassification risk they insure, and how these contract characteristics depend on a consumer's income profile. We then turn in Sections 6 and 7 to welfare analysis: In Section 6 we analyze the extent to which these optimal dynamic contracts eliminate the welfare losses from reclassification risk, and how this depends on consumers' income profiles, and in Section 7 we compare the welfare achieved by these contracts to that in an ACA-like exchange.

## 5.1 Front-loading and Reclassification Risk

We begin by considering the optimal contract for a consumer with flat net income (corresponding to the solid curve in Figure 3.4, an income profile that creates no borrowing or savings motive when a consumer faces the at-birth ex ante expected medical expenses at each age).

In our context, a contract specifies a premium, or equivalently, a consumption level for each possible history of states at each age from 25 to 65. There are too many histories and concurrent premiums/consumption levels to present: instead we focus on select attributes of the contract. First, we look in detail at the early contract periods, which provide intuition for the form of the equilibrium contract over longer time horizons. The first-period premiums, consumption levels, actuarial costs, and front-loading are presented in Table 7 for consumers whose health status at ages 24 and 25 were the same (i.e., with $\lambda_{24} = \lambda_{25}$). In all but the worst state ($\Lambda_{25} = (7, 7)$), premiums are larger than actuarial costs: consumers front-load premiums to transfer utility to future states with negative health shocks. For example, for

[29] These probabilities are computed using $C_{t+1}^\theta$ and the transitions $\overline{f}(\cdot)$.

| First-Year Equilibrium Contract Terms: Flat Net Income | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | $\lambda_{24} = \lambda_{25}$ | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Premium | 2.294 | 3.472 | 4.156 | 6.255 | 11.444 | 13.733 | 20.511 |
| First-Year Costs | 0.837 | 1.375 | 1.973 | 3.054 | 4.358 | 6.842 | 20.511 |
| Front-Loading | 1.457 | 2.098 | 2.182 | 3.201 | 7.086 | 6.891 | 0 |
| Consumption | 54.765 | 53.586 | 52.903 | 50.803 | 45.615 | 43.326 | 36.548 |

Table 7: First-year contract terms in the equilibrium long-run contract for men with a flat net income path, showing first-year premiums, expected costs, the extent of front-loading, and consumption levels (thousands of dollars).

the healthiest consumer at the beginning of the year ($\Lambda_{25} = (1, 1)$), the premium is \$2,294 despite average costs of only \$837.

The extent of front-loading rises as the consumer's health state worsens to bin 5 out of 7, and then declines to zero for the sickest bin of consumers (7 out of 7). The extent of front-loading depends on both the current state and also on the implications of the current state for future health. While the healthiest type can afford the most front-loading, he might benefit the least. This is why maximum front-loading occurs for consumers in the middle of the ex ante health range, rather than for the healthiest consumers. The least healthy, on the other hand, have very high current costs (and, hence, high marginal utility of consumption) compared to their expected future costs (recall Table 5); for them, front-loading isn't worthwhile.

Tables 8 and 9 present second-period premiums and consumptions, respectively, for the seven possible age-26 health statuses that can follow the seven age-25 health states considered in Table 7. (These age-26 health status realizations $\lambda_{26}$ give rise to the age-26 health state $\Lambda_{26} = (\lambda_{25}, \lambda_{26})$.) Certain patterns are indicative of the longer-run structure of the contract. First, second-period premiums and consumptions display extensive pooling which takes place in states for which the lapsation constraint is not binding. For example, if a consumer was in the healthiest possible state at age 25, $\Lambda_{25} = (1, 1)$, all second-year states $\Lambda_{26} = (1, \lambda_{26})$ with $\lambda_{26} > 1$ have the same consumption of \$54,765, an amount equal to his first-year consumption. The lapsation constraint does not bind for this consumer when $\lambda_{26} > 1$ because the first period front-loaded amount suffices to make outside offers less attractive than the current consumption guarantee. Only when $\lambda_{26} = 1$ does the lapsation constraint bind for this consumer, resulting in an increase in consumption to \$54,791 (and a corresponding reduction in the premium).

The lapsation constraint binds for more and more second-year states the sicker the consumer was at the start of the contract. For consumers initially in the sickest health state, $\Lambda_{25} = (7, 7)$, all age-26 health states involve different consumption levels that also differ from

| Second-Year Equilibrium Premiums: Flat Net Income | | | | | | | | First-Year |
|---|---|---|---|---|---|---|---|---|
| | | | | $\lambda_{26}$ | | | | Premium |
| $\lambda_{24} = \lambda_{25}$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
| 1 | 2.403 | 2.429 | 2.429 | 2.429 | 2.429 | 2.429 | 2.429 | 2.294 |
| 2 | 2.864 | 3.445 | 3.607 | 3.607 | 3.607 | 3.607 | 3.607 | 3.472 |
| 3 | 2.941 | 3.716 | 4.088 | 4.291 | 4.291 | 4.291 | 4.291 | 4.156 |
| 4 | 2.470 | 3.063 | 4.454 | 6.004 | 6.390 | 6.390 | 6.390 | 6.255 |
| 5 | 2.397 | 2.891 | 3.850 | 6.224 | 11.240 | 11.578 | 11.578 | 11.444 |
| 6 | 2.349 | 2.667 | 3.184 | 5.669 | 11.385 | 13.599 | 13.867 | 13.733 |
| 7 | 2.385 | 2.470 | 2.650 | 4.975 | 15.809 | 14.798 | 20.511 | 20.511 |

Table 8: First- and second-year premiums 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 Consumptions: Flat Net Income | | | | | | | | First-Year |
|---|---|---|---|---|---|---|---|---|
| | | | $\lambda_{26}$ | | | | | 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 |
| 2 | 54.330 | 53.749 | 53.586 | 53.586 | 53.586 | 53.586 | 53.586 | 53.586 |
| 3 | 54.253 | 53.478 | 53.106 | 52.903 | 52.903 | 52.903 | 52.903 | 52.903 |
| 4 | 54.724 | 54.13 | 52.739 | 51.189 | 50.803 | 50.803 | 50.803 | 50.803 |
| 5 | 54.796 | 54.303 | 53.343 | 50.970 | 45.954 | 45.615 | 45.615 | 45.615 |
| 6 | 54.845 | 54.527 | 54.010 | 51.524 | 45.809 | 43.595 | 43.326 | 43.326 |
| 7 | 54.809 | 54.724 | 54.544 | 52.219 | 41.384 | 42.396 | 36.683 | 36.548 |

Table 9: 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).

the first-period consumption level: long-run contracts cannot provide any insurance against reclassification risk in year 2 for this consumer as his first-year needs were so great as to preclude any front loading. For this consumer, the long-run contract continuation at age 26 simply matches the best contract he could get on the market given his age-26 health state.

## 5.2 Effects of Income profiles

The equilibrium contracts offered depend crucially on a consumer's rate of income growth over his lifetime. When income is relatively low early in life, and hence the marginal utility of consumption is relatively high, front-loading is quite costly for utility.

| First-Year Equilibrium Contract Term: Downscaled Managers | | | | | | | |
|---|---|---|---|---|---|---|---|
| | $\lambda_{24} = \lambda_{25}$ | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Premium | 1.165 | 2.196 | 2.740 | 3.487 | 8.089 | 10.887 | 20.511 |
| First-Year Costs | 0.837 | 1.375 | 1.973 | 3.054 | 4.358 | 6.842 | 20.511 |
| Front-Loading | 0.328 | 0.821 | 0.767 | 0.433 | 3.732 | 4.046 | 0 |
| Consumption | 32.520 | 31.490 | 30.945 | 30.199 | 25.596 | 22.799 | 13.175 |

Table 10: First-year contract terms in the equilibrium long-run contract for men with a downscaled manager income path, showing first-year premiums, expected costs, the extent of front-loading, and consumption levels (thousands of dollars).

Table 10 presents first-period (age-25) contract characteristics for "downscaled managers," and is the analog to Table 7 for flat net income. Recall that, as shown in Figure 3.4, a downscaled manager income profile proportionally scales down the income of a manager to match the net present value of a non-manager's lifetime income profile. The table makes clear that for downscaled managers, the extent of front-loading is much more limited than in the flat net income case, which translates into less generous consumption guarantees later in life. For example, a downscaled manager who is in the healthiest state ($\Lambda_{25} = (1, 1)$) at age 25 front-loads only \$328, compared to \$1,457 for a consumer with flat net income. Essentially, the rapidly rising income makes paying extra early in life for long-term insurance quite costly, as marginal utility is high early compared to what is expected later in life.

Tables 11 and 12 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.

# 6 Results: Welfare

We now turn to the welfare analysis of these dynamic contracts. We measure and compare the welfare they achieve to several alternatives. For each market setup and potential income profile considered, we compute a lifetime certainty equivalent. The certainly equivalent represents the constant consumption for the forty years of life from age 25 to 65 that makes the consumer as well off as in a given market setup. Specifically, we compare the certainty equivalent of optimal dynamic contracts with one-sided commitment, denoted by $CE_D$, to the three benchmarks we have described previously (see Section 2.1 for formal definitions):[30]

[30] In Section 7 we compare dynamic contracts as well to an ACA-style insurance exchange.

| Second-Year Equilibrium Premiums: Downscaled Managers | | | | | | | | First-Year |
|---|---|---|---|---|---|---|---|---|
| | | | | $\lambda_{26}$ | | | | 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 |
| 2 | 1.542 | 2.190 | 2.356 | 3.414 | 4.042 | 4.042 | 4.042 | 2.196 |
| 3 | 1.636 | 2.077 | 2.755 | 4.587 | 4.587 | 4.587 | 4.587 | 2.740 |
| 4 | 0.884 | 1.405 | 2.548 | 3.490 | 5.334 | 5.334 | 5.334 | 3.487 |
| 5 | 0.841 | 1.375 | 2.249 | 3.608 | 8.066 | 9.936 | 9.936 | 8.089 |
| 6 | 0.837 | 1.375 | 1.973 | 3.320 | 8.219 | 10.882 | 12.734 | 10.887 |
| 7 | 0.843 | 1.375 | 1.973 | 3.054 | 11.895 | 11.851 | 20.511 | 20.511 |

Table 11: 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).

| Second-Year Equilibrium Consumption: Downscaled Managers | | | | | | | | First-Year |
|---|---|---|---|---|---|---|---|---|
| | | | | $\lambda_{26}$ | | | | 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 |
| 2 | 33.990 | 33.342 | 33.176 | 32.118 | 31.490 | 31.490 | 31.490 | 31.490 |
| 3 | 33.896 | 33.455 | 32.777 | 30.945 | 30.945 | 30.945 | 30.945 | 30.945 |
| 4 | 34.648 | 34.127 | 32.984 | 32.042 | 30.199 | 30.199 | 30.199 | 30.199 |
| 5 | 34.691 | 34.158 | 33.283 | 31.925 | 27.466 | 25.596 | 25.596 | 25.596 |
| 6 | 34.695 | 34.158 | 33.559 | 32.212 | 27.314 | 24.651 | 22.799 | 22.799 |
| 7 | 34.689 | 34.158 | 33.559 | 32.478 | 23.637 | 23.681 | 15.021 | 13.175 |

Table 12: 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).

**(i)** The first-best, fully-smoothed consumption $C^*$, which equates the marginal utility of consumption across periods and states. This is the welfare achievable were long-term contracts with two-sided commitment feasible;

**(ii)** The certainty equivalent from spot contracts that fully insure event risk in every period and state, but leave reclassification risk across periods fully uninsured, denoted by $CE_{SPOT}$;

**(iii)** The constant consumption equivalent of the No Borrowing/No Saving constrained first best, in which risk is fully insured in each period but neither borrowing nor saving is possible, denoted $C^*_{NBNS}$.

## 6.1 Welfare Effects Conditional on a Consumer's Age-25 Health State

Table 13 shows welfare outcomes for Utah men with a flat net income profile who arrive at age 25 in each of the seven health states $\Lambda_{25}$ in which their age-24 and age-25 health status is the same (i.e., in which $\lambda_{24} = \lambda_{25}$; these are the health states examined in Tables 3-12). For each age-25 health state, column (1) reports the annual consumption level $C^*$ in a first-best contract that starts at age 25 given the income profile and expected future medical expenses the consumer faces given his age-25 health state.[31] It ranges from \$54,960 for the healthiest consumer state $\Lambda_{25} = (1,1)$ to \$49,330 for a consumer in the worst state $\Lambda_{25} = (7,7)$. Column (2) shows $C^*_{NBNS}$, the constant consumption equivalent of the constrained first-best outcome that does not allow for intertemporal consumption smoothing.

For consumers with rising net income, $C^*_{NBNS}$ may be a more relevant benchmark of the losses from spot contracting and of how well optimal dynamic contracts with one-sided commitment do at eliminating reclassification risk, since saving and borrowing on their own can greatly improve utility for steep net income profiles. (For a healthy consumer with flat net income, however, this certainty equivalent is very close to $C^*$.) Column (3) shows welfare under spot contracts for each of these consumers, while Column (5) shows the welfare loss from spot contracting relative to this benchmark, $\frac{C^*_{NBNS} - CE_{SPOT}}{C^*_{NBNS}}$, a measure that captures solely the loss under spot contracting arising from reclassification risk. This welfare loss is very large: many of these consumers lose roughly 18% of their *lifetime* (age 25-65) certainty equivalent because of reclassification risk.

[31] In this and the other three tables in this subsection, "ex ante" certainty equivalents are calculated from the perspective of a consumer who arrives at age 25 in a particular health state. Thus, for example, the first-best consumption of a consumer with flat net income will differ across consumers with different health states $\Lambda_{25}$ because of their differing expected medical costs.

| Initial health | | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|---|
| $\lambda_{24}$ | $\lambda_{25}$ | $C^*$ | $C^*_{NBNS}$ | $CE_{SPOT}$ | $CE_D$ | $\frac{C^*_{NBNS}-CE_{SPOT}}{C^*_{NBNS}}$ | $\frac{CE_D-CE_{SPOT}}{C^*_{NBNS}-CE_{SPOT}}$ | $\frac{CE_D-CE_{SPOT}}{C^*-CE_{SPOT}}$ |
| 1 | 1 | 55.14 | 55.10 | 45.31 | 55.03 | 0.178 | 0.994 | 0.989 |
| 2 | 2 | 54.96 | 54.94 | 45.03 | 54.80 | 0.180 | 0.986 | 0.984 |
| 3 | 3 | 54.84 | 54.83 | 44.92 | 54.62 | 0.181 | 0.979 | 0.978 |
| 4 | 4 | 54.36 | 54.34 | 44.45 | 53.69 | 0.182 | 0.934 | 0.933 |
| 5 | 5 | 52.86 | 52.14 | 42.47 | 49.62 | 0.185 | 0.739 | 0.688 |
| 6 | 6 | 51.51 | 49.42 | 41.53 | 47.08 | 0.160 | 0.703 | 0.556 |
| 7 | 7 | 49.33 | 42.61 | 39.94 | 40.72 | 0.063 | 0.291 | 0.083 |

Table 13: Long-run welfare results showing the certainty equivalent annual consumption of different insurance institutions under various initial health states, the flat net 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.

Column (4) presents the certainty equivalent for dynamic contracts with one-sided commitment, $CE_D$. As expected $CE_D$ lies between $C^*_{NBNS}$ and $CE_{SPOT}$. Column (5) shows the fraction of the welfare gap between the No-borrowing/No-saving constrained first-best and spot contracts that these dynamic contracts close, $\frac{CE_D-CE_{SPOT}}{C^*_{NBNS}-CE_{SPOT}}$. Overall, dynamic contracts are extremely effective at reducing reclassification risk for consumers who arrive at age 25 in excellent health: for a consumer in age-25 health state $\Lambda_{25} = (1, 1)$, dynamic contracts close 99.4% of the gap between $C^*_{NBNS}$ and $C_{SPOT}$. But they are very ineffective for consumers who arrive at age 25 in poor health. At the extreme, a consumer who arrives at age 25 in the worst health state, $\Lambda_{25} = (1, 1)$, dynamic contracts recover only 29.1% of the welfare loss due to reclassification risk under spot contracting. The reason for this pattern is that consumers who arrive at age 25 in poor health have a high level of current medical expenses, which makes front-loading very costly and therefore greatly limits the effectiveness of dynamic contracts. For completeness, column (7) reports $\frac{CE_D-CE_{SPOT}}{C^*-CE_{SPOT}}$, which shows the proportion of the gap between spot contracting and the first-best closed by dynamic contracts.

Tables 14–16 report the same welfare statistics but for consumers with non-manager, manager, and downscaled manager income profiles. Welfare losses from reclassification risk (shown, again, in column (5)) are similarly large for consumers with rising income profiles. Comparing column (6) for the cases with rising income profiles to the flat net income profile case shows that rising income profiles reduce the effectiveness of dynamic contracts by making front-loading more costly, since with a rising income profile the marginal utility of a current dollar is larger than the marginal utility of future dollars. The effect is particularly dramatic for consumers in poor health states: for example, for a consumer in state $\Lambda_{25} = (7, 7)$,

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

Table 14: 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.

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

Table 15: 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.

dynamic contracts reduce the loss from reclassification risk by 29.1% if the consumer has a flat net income profile, but by only 7.6% if he has a downscaled manager income profile.

## 6.2 Welfare Effects from the Perspective of an Unborn Consumer

We now turn to evaluating the welfare effects of dynamic contracts from the perspective of an unborn consumer who does not know what his age-25 health state will be. We compute these welfare measures using the distribution of age-25 health states that we see empirically among our age-25 Utah men. Table 17 shows the results.

Recall that, to ease comparisons across different income profiles, all profiles we consider have the same net present value of income, except for the manager's profile when it is not downscaled. Column (1) of Table 17 shows the first-best consumption (reflecting the present

| Initial health | | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|---|
| $\lambda_{24}$ | $\lambda_{25}$ | $C^*$ | $C^*_{NBNS}$ | $CE_{SPOT}$ | $CE_D$ | $\frac{C^*_{NBNS}-CE_{SPOT}}{C^*_{NBNS}}$ | $\frac{CE_D-CE_{SPOT}}{C^*_{NBNS}-CE_{SPOT}}$ | $\frac{CE_D-CE_{SPOT}}{C^*-CE_{SPOT}}$ |
| 1 | 1 | 55.14 | 39.07 | 32.73 | 38.76 | 0.162 | 0.951 | 0.269 |
| 2 | 2 | 54.96 | 38.47 | 30.75 | 37.87 | 0.201 | 0.922 | 0.294 |
| 3 | 3 | 54.84 | 37.94 | 30.19 | 37.26 | 0.204 | 0.912 | 0.287 |
| 4 | 4 | 54.36 | 37.07 | 32.40 | 36.13 | 0.126 | 0.799 | 0.170 |
| 5 | 5 | 52.86 | 34.23 | 25.87 | 31.31 | 0.244 | 0.651 | 0.202 |
| 6 | 6 | 51.51 | 30.82 | 24.26 | 28.47 | 0.213 | 0.642 | 0.155 |
| 7 | 7 | 49.33 | 20.56 | 19.89 | 19.94 | 0.032 | 0.076 | 0.002 |

Table 16: Long-run welfare results showing the certainty equivalent annual consumption of different insurance institutions under various initial health states, the downscaled 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.

| | (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 | 54.67 | 54.67 | 44.35 | 48.83 | 0.189 | 0.434 | 0.434 |
| Non-mngr | 54.67 | 47.37 | 36.96 | 38.08 | 0.220 | 0.108 | 0.063 |
| Manager | 85.00 | 55.67 | 45.44 | 45.91 | 0.184 | 0.046 | 0.012 |
| Downs Mngr | 54.67 | 37.68 | 27.35 | 28.13 | 0.274 | 0.075 | 0.028 |

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

value of income minus expected medical expenses, now calculated from the perspective of a consumer who does not yet know their age-25 health state), which are \$54,670 for the flat-net, non-manager, and downscaled manager income profiles, and \$85,000 for the manager income profile. Column (2) gives the certainty equivalent of the No-Borrowing/No-Saving full insurance regime, again based on expected medical expenses calculated from the perspective of a consumer who does not yet know their age-25 health state.

Column (3) shows welfare outcomes under spot contracts, with no protection for reclassification risk, while column (4) shows the certainty equivalent of the optimal dynamic contract regime, both from the perspective of an unborn consumer.[32] The welfare loss from reclassification risk, shown in column (5), ranges between 18.4% and 27.4% of lifetime certainty

[32] $C^*_{NBNS}$ in Table 17 equals the expectation of the $C^*_{NBNS}$ values for consumers in the various age-25 health states, evaluated using the empirical distribution of age-25 health states. In contrast, because of risk aversion, $CE_D$ in Table 17 is less than the expectation of the $CE_D$ values for consumers in the various age-25 health states.

equivalent (relative to the no-borrowing/no-savings benchmark).[33] For a flat net income profile, column (6) shows that 43.4% of this gap is recovered by dynamic contracts. In contrast, since consumers with steeper income profiles dislike front-loading, dynamic contracts recover a very small portion of this welfare loss for these consumers (7.5% for downscaled managers).

In summary, from the perspective of an unborn Utah male with a flat net income profile, optimal dynamic contracts would be moderately effective at reducing the reclassification risk they face. However, from the perspective of an unborn Utah male with a rising net income profile, these contracts would not be very effective. The moderate effectiveness with a flat net income reflects the fact that dynamic contracts poor performance for consumers who arrive at age 25 in poor health, and thus with high immediate consumption needs, has a substantial effect on the value of dynamic contracts from the ex ante perspective of an unborn consumer. Once coupled with more steeply rising income paths, which further accentuate the value of net income when young, dynamic contracts prove rather ineffective from an ex ante welfare perspective.

## 6.3 Insurance of Pre-age-25 Health Risk

The results in Tables 13-16 indicate that in a regime in which consumers sign long-term dynamic contracts upon reaching age 25, their post-age-25 welfare is greatly affected by the realization of their age-25 health state. In Table 18 we examine how welfare would be affected if the government insured consumers' pre-age-25 health realizations. We consider this in two ways. First, we ask what the expected per capita cost would be for the government to ensure that each consumer's continuation certainty equivalent starting at age 25 is the same as if he had reached age 25 in the healthiest state, $\Lambda_{25} = (1, 1)$. Second, we derive the set of break-even subsidies that most efficiently insure the age-25 health risk these consumers face. This involves finding the subsidy or tax for consumers in each of the 49 health age-25 states such that (i) the government breaks even in expectation, and (ii) the welfare of an unborn consumer is maximized.[34]

Column (5) reports the (one-time) expected per capita cost of subsidies that insure that, under a regime of optimal dynamic contracts that begin at age 25, every consumer has the same certainty equivalent as if he had arrived at age 25 in health state $\Lambda_{25} = (1, 1)$. This cost ranges from \$5,470 for a manger income profile to \$12,050 for a flat net income

[33] In Section 8.2 we extend this analysis to allow precautionary savings in the spot contracting regime and find similar results. This is not surprising as the main loss with a rising income path from a lack of interporal smoothing comes from the inability to borrow.

[34] Note that this does not generally yield equal certainty equivalents for the 49 health states because the marginal utility of a dollar subsidy is state-dependent.

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Income profile | $C^*$ | $C^*_{NBNS}$ | $CE_{SPOT}$ | $CE_D$ | Required subsidy | $CE_{D+}$ | $\frac{CE_{D+}-CE_{SPOT}}{C^*_{NBNS}-CE_{SPOT}}$ |
| Flat net | 54.67 | 54.67 | 44.35 | 48.83 | 12.05 | 53.91 | 0.926 |
| Non-mngr | 54.67 | 47.37 | 36.96 | 38.08 | 8.81 | 45.24 | 0.795 |
| Manager | 85.00 | 55.67 | 45.44 | 45.91 | 5.47 | 54.05 | 0.842 |
| Downs Mngr | 54.67 | 37.68 | 27.35 | 28.13 | 7.07 | 35.79 | 0.817 |

Table 18: Unborn consumer welfare results showing the certainty equivalent annual consumption of different insurance institutions under various income profiles, a discount factor of 0.975, and constant absolute risk aversion equal to 0.0004. Column (5) shows the expected one-time subsidy required at age 25 for the consumer to have in all age-25 states the same level of welfare as if he had been in the healthiest possible age-25 health state. Column (6) shows the certainty equivalent welfare level resulting from a balanced budget scheme that optimally insures the consumer's pre-age-25 health risk prior to the start of dynamic contracting at age 25. Units in columns (1)-(6) are 1000s of dollars.

profile. Column (6) reports the certainty equivalent achieved instead when optimal dynamic contracts starting at age 25 are coupled with a break-even government insurance scheme that insures against consumers' pre-age-25 health risks, which we label as $CE_{D+}$, while column (7) shows what fraction of the reclassification risk losses are recovered through a combination of this kind of government policy and optimal dynamic contracts starting at age 25. When combined with this pre-age-25 insurance, dynamic contracts eliminate roughly 80-90% of reclassification risk. One can think of these policies to insure age-25 risk as similar in spirit to the risk-adjustment, risk-corridor, and reinsurance regulations present in the ACA (and many other current environments) but applied to age-25 consumers choosing dynamic contracts instead of consumers of all ages choosing year-to-year contracts. The results show that long-term dynamic contracts can be rather effective at eliminating reclassification risk in combination with government insurance of consumers' health risks that occur before they reach the insurance market at age 25.

# 7 Comparison to the ACA

One of the most significant features of the health insurance exchanges created by the ACA was their ban on the pricing of pre-existing conditions. In this section, we examine how dynamic contracts do at eliminating reclassification risk compared to an ACA-like exchange. As discussed in Handel, Hendel, and Whinston (2015), while the ACA fully eliminated reclassification risk, it created adverse selection, as consumers with differing health could not be differentially priced. This adverse selection led to significant unraveling in the exchange model studied in Handel, Hendel, and Whinston (2015), so that in most cases all consumers ended up obtaining insurance contracts covering only 60% of their expenses.

In this section we compare the welfare achievable with optimal dynamic contracts, both with and without government insurance of pre-age-25 health risk, to the level of welfare that would arise in an ACA-like insurance exchange.[35] Because consumers end up only partly insured under the ACA, computing welfare requires as an input the full distribution of health expenses conditional on an individual's health state, rather than just its mean. We have previously estimated this full distribution for the consumers in the Handel, Hendel, and Whinston (2015) sample and we make use of this information here. Specifically, we examine how dynamic contracts would perform for the Handel, Hendel, and Whinston (2015) consumers if they had the transitions that we have estimated for Utah men, and compare it to what the ACA would achieve.[36]

We compute welfare under ACA-like regulation imposing (i) one-year contracts, (ii) community rating (no health-state based pricing allowed), (iii) age-based pricing, (iv) a fully enforced mandate, requiring insurance purchase, and (v) insurers that offer plans covering specific actuarial values, with a minimum plan covering 60% of an average individual's spending that the market unravels to. We denote the certainty equivalent of this outcome by $CE_{ACA}$.

Columns (1)-(4) of Table 19 give the (unborn) certainty equivalents for institutions that were in Table 17[37], while column (5) gives the corresponding certainty equivalent under the ACA. Comparing columns (3) and (5), dynamic contracts without government insurance of pre-age-25 health risk are worse than the ACA for all income profiles. One possible reason for this difference is that the ACA's community rating implicitly insures consumers' pre-age-25 health risk.

Column (4), on the other hand, shows the welfare level achievable for this sample when dynamic contracts are coupled with insurance of pre-age-25 health risk. Columns (6) and (7) show how much of the gap between spot contracting and the No-Borrowing/No-Savings

[35] The ACA insured pre-age-25 risk through its ban on pre-existing conditions, so a natural comparison is to dynamic contracts combined with insurance of pre-age-25 risk.

[36] The sample in Handel, Hendel, and Whinston (2015) is too small to estimate second-order Markov transitions, while for the Utah sample we have only mean health costs conditional on health states, not the full distributions. For these reasons we combine the data in the two samples for this analysis. We have examined the effect of dynamic contracts in both the Utah and Handel, Hendel, and Whinston (2015) samples with first-order Markov processes and found similar results. By way of comparison to the male Utah sample, in the Handel, Hendel, and Whinston (2015) sample, the expected costs conditional on each of the seven health statuses (going from healthiest to sickest) are \$1,131, \$2,290, \$3,780, \$3,975, \$5,850, \$10,655, and \$18,554. Comparing to Table 2, Utah expenses were lower in bins 1-6, but higher for consumers with the worst health status.

[37] Note that the welfare levels for $C^*$ and $C^*_{NBNS}$ here differ from those in Table 17 because mean health expenses in our sample of Utah men differ from those in the Handel, Hendel, and Whinston (2015) sample, as noted in the previous footnote.

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Income profile | $C^*$ | $C^*_{NBNS}$ | $CE_D$ | $CE_{D+}$ | $CE_{ACA}$ | $\frac{CE_{D+}-CE_{SPOT}}{C^*_{NBNS}-CE_{SPOT}}$ | $\frac{CE_{ACA}-CE_{SPOT}}{C^*_{NBNS}-CE_{SPOT}}$ |
| Flat net | 54.62 | 54.62 | 52.70 | 54.12 | 52.90 | 0.890 | 0.622 |
| Non-mngr | 54.62 | 47.25 | 43.19 | 45.90 | 46.44 | 0.710 | 0.826 |
| Manager | 84.95 | 55.52 | 51.25 | 54.52 | 54.77 | 0.777 | 0.833 |
| Downs Mngr | 54.62 | 37.55 | 33.37 | 36.36 | 36.78 | 0.740 | 0.832 |

Table 19: Long-run welfare results showing the certainty equivalent annual consumption of different insurance institutions, including the ACA, for the large employer sample of Handel, Hendel, and Whinston (2015). Assumes a discount factor of 0.975 for consumers with median estimated constant absolute risk aversion equal to 0.0004. Units in columns (1)-(5) are 1000s of dollars.

benchmark is closed by dynamic contracts with insurance of pre-age-25 health risk and the ACA, respectively.[38] Comparing columns (6) and (7) in Table 19 reveals that the ACA environment is preferred to the dynamic contracting environment even with insurance of pre-age-25 health risk for non-managers, managers, and downscaled managers, whose incomes rise over time. For these individuals, the desire to front-load when young and income is relatively low is limited, which reduces the benefits from dynamic contracts. In contrast, individuals with flat net income profiles prefer dynamic contracts with insurance of pre-age-25 health risks to the ACA environment.

# 8 Extensions

In this section we consider several extensions. First, we consider how the presence of switching costs would affect the gains from dynamic contracts. Second, we study how the ability to engage in precautionary savings would affect our conclusions. Third, motivated by the concern that risk aversion could differ from typical estimates in the health insurance literature for the large losses created by reclassification risk, we examine the welfare effects of dynamic contracts for lower levels of risk aversion than the level we have considered above.

[38] The No-Borrowing/No-Savings outcome is the same as the ACA outcome except that it provides 100% coverage rather than unraveling to 60%; the difference between $C^*_{NBNS}$ and $CE_{ACA}$ therefore reflects the cost of adverse selection under the ACA. Table 19 shows that the cost of adverse selection ranges from roughly \$750 per year for managers, downscaled managers, and non-managers to roughly \$1700 per year for consumer with flat net income profiles.

| | | Second-year consumption | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | $\lambda_{26}$ | | | | First-year |
| $\lambda_{24}$ | $\lambda_{25}$ | 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 |
| 2 | 2 | 54.305 | 53.986 | 53.986 | 53.986 | 53.986 | 53.986 | 53.986 | 53.986 |
| 3 | 3 | 54.227 | 53.453 | 53.340 | 53.340 | 53.340 | 53.340 | 53.340 | 53.340 |
| 4 | 4 | 54.698 | 54.105 | 52.714 | 51.214 | 51.214 | 51.214 | 51.214 | 51.214 |
| 5 | 5 | 54.770 | 54.277 | 53.318 | 50.945 | 45.947 | 45.947 | 45.947 | 45.947 |
| 6 | 6 | 54.819 | 54.502 | 53.985 | 51.499 | 45.784 | 43.661 | 43.661 | 43.661 |
| 7 | 7 | 54.783 | 54.699 | 54.519 | 52.194 | 41.359 | 42.371 | 36.996 | 36.996 |

Table 20: 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.

## 8.1 Switching Costs

Recent evidence from health insurance markets [Handel (2013), Ho et al. (2016)] points to substantial inertia in insurance choice. Switching costs have the potential to improve how dynamic contracts perform. Our basic model assumes that consumers lapse whenever they get a better offer. As discussed in Section 2.4, switching costs relax the lapsation constraints, which can enhance commitment and the welfare achievable with optimal dynamic contracts. (Introducing switching costs can also be viewed as allowing for imperfect access to credit markets, since switching costs enable debt contracts by making consumers willing to repay their debts rather than switch to another lender.)

Table 20 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 8, 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.

Table 21 shows the ex ante (at-birth) welfare achieved by dynamic contracts for different levels of switching costs and our four income profiles in the male Utah sample. As expected, welfare is monotonic in the switching cost. Qualitatively, as switching costs increase from zero to infinity, welfare in the optimal dynamic contract with one-sided commitment approaches the first-best (two-sided commitment) level.

Notice that it takes extremely large switching costs to achieve welfare close to first best for consumers with steeply rising manager and downscaled-manager income profiles. The reason is that consumption smoothing through borrowing requires a lot of commitment, especially

| Switching cost | Flat net | Non-manager | Manager | Downs mngr |
|---|---|---|---|---|
| 0 | 48.83 | 38.08 | 45.91 | 28.13 |
| 0.1 | 48.85 | 38.13 | 45.98 | 28.19 |
| 1 | 49.02 | 38.54 | 46.60 | 28.69 |
| 5 | 49.67 | 39.89 | 48.63 | 30.31 |
| 10 | 50.37 | 41.13 | 50.46 | 31.79 |
| 50 | 53.37 | 47.06 | 58.95 | 38.92 |
| 100 | 54.15 | 51.47 | 65.58 | 44.57 |
| 500 | 54.19 | 54.19 | 84.51 | 54.19 |
| $C^*$ | 54.67 | 54.67 | 85.00 | 54.67 |
| $C^*_{NBNS}$ | 54.67 | 47.37 | 55.67 | 37.68 |

Table 21: At-birth welfare (in \$1,000s) for Utah men from optimal dynamic contracts ($CE_D$) under discount factor of 0.975 and risk aversion of 0.0004 for different levels of switching costs and four income profiles.

when the income profile is steep. Thus, an extremely large switching cost is necessary to achieve the first best. Somewhat more moderate switching costs deliver welfare close to the no-borrowing/no-saving benchmark.

Using the same hybrid sample as in Section 6, Table 22 reports for each income profile the switching cost that is needed to achieve the same level of welfare as the ACA. For a flat net income profile, switching costs of \$1,970 suffice, while for rising income profiles switching costs between \$7,680-\$10,470 make dynamic contracts (without any pre-age-25 insurance) as good as the ACA. Handel (2013) estimates a mean switching cost of \$2,032 in a static model of choice for the same population as in our Handel, Hendel, and Whinston (2015) sample. Since the dynamic gain from switching likely extends over multiple periods, the comparable value for our model is likely significantly higher than this \$2,032, and so switching costs may well be in the range that make dynamic contracts (without insurance of pre-age-25 health risk and with our baseline risk aversion) preferable to the ACA for some income profiles.[39,40]

## 8.2 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.

[39] Illanes (2017), for example, estimates a lower bound switching cost of \$1200 in a dynamic model of choice for the Chilean pension market; he shows that the estimate from a static model of choice in his sample is \$117.

[40] Note, however, that our ACA model assumes that there are no switching costs; modeling insurance exchange competition with switching costs remains an open issue.

| Income profile | Switching costs |
|---|---|
| Flat net | 1.97 |
| Non-mngr | 13.59 |
| Manager | 7.68 |
| Downs Mngr | 10.47 |

Table 22: Switching costs required to equalize certainty equivalents from optimal dynamic contracts and the ACA insurance regime (in thousands of dollars). Computed using the hybrid sample of Section 6, constant absolute risk aversion equal to 0.0004, and a discount factor of 0.975.

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 with the different health states.[41] 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 23 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.

[41] 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.

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Income profile | $C^*_{NBNS}$ | $CE_{SPOT}$ | $CE_{SPOTwS}$ | $CE_D$ | $\frac{C^*_{NBNS}-CE_{SPOTwS}}{C^*_{NBNS}}$ | $\frac{CE_{SPOTwS}-CE_{SPOT}}{C^*_{NBNS}-CE_{SPOT}}$ | $\frac{CE_D-CE_{SPOTwS}}{C^*_{NBNS}-CE_{SPOTwS}}$ |
| 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 23: 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 $CE_{SPOTwS}$.

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 downscaled managers.

## 8.3 Risk Aversion

So far our analysis has used the risk preferences estimated in Handel, Hendel, and Whinston (2015). We now consider the robustness of the analysis with respect to the degree of risk aversion. We are particularly interested in lower risk aversion. The reason is that we are afraid our estimate of risk preferences, estimated from choices among health insurance contracts with out-of-pocket caps, might not reflect consumers' risk tolerance for the larger stakes associated with reclassification risk.[42]

Table 24 presents the welfare comparisons for risk aversion of 0.00008, five times lower than that in our main analysis. For a consumer with $50,000 of consumption, this corresponds to a CRRA risk aversion coefficient of 4, roughly the level suggested in the macro literature on consumption disasters (e.g., Barro (2006)). To put the coefficients in perspective, consider a lottery that assigns the costs associated with each of the seven health statuses, with each having equal probability. For the costs we used in Section 7, our 0.0004 risk aversion coefficient estimate implies a willingness to pay of $7,222 to avoid the uncertainty associated with this risky prospect. Instead, the lower risk aversion coefficient leads to a willingness to pay of $1,491.

[42] See Rabin (2000) for a discussion of issues with CARA and scaling of gambles, and Collier et al (2017) for evidence of differing risk aversion for small versus large stakes decisions.

| Income profile | (1) $C^*$ | (2) $CE_{SPOT}$ | (3) $CE_D$ | (4) $C^*_{NBNS}$ | (5) $\frac{C^*_{NBNS}-CE_{SPOT}}{C^*_{NBNS}}$ | (6) $\frac{CE_D-CE_{SPOT}}{C^*_{NBNS}-CE_{SPOT}}$ | (7) $\frac{CE_D-CE_{SPOT}}{C^*-CE_{SPOT}}$ |
|---|---|---|---|---|---|---|---|
| Flat net | 54.67 | 52.99 | 54.38 | 54.67 | 0.031 | 0.827 | 0.827 |
| Non-mngr | 54.67 | 51.22 | 52.05 | 52.72 | 0.028 | 0.550 | 0.239 |
| Manager | 85.00 | 68.45 | 68.78 | 69.73 | 0.018 | 0.256 | 0.020 |
| Downs Mngr | 54.67 | 46.44 | 46.95 | 47.81 | 0.029 | 0.372 | 0.062 |

Table 24: Long-run welfare results for Utah men showing the certainty equivalent consumption of different insurance institutions for a discount factor of 0.975 and constant absolute risk aversion equal to 0.00008.

Lowering risk aversion substantially reduces the loss associated with reclassification risk (captured by the gap between $C^*_{NBNS}$ and $CE_{SPOT}$). The loss is between 1.8% and 3.1%, depending on the income profile. For the higher 0.0004 risk aversion, the loss was between 18.4% and 27.4%. Still, a loss of 1-4% of age 25-65 certainty equivalent is not insubstantial.

While the loss from reclassification risk is lower, the reductions in reclassification risk from long-term contracting, as captured in column (6) by how much of the gap between the welfare under optimal dynamic contracts and that under the No-Borrowing/No-Saving benchmark, are much larger, ranging between 25.6% for the manager income profile and 82.7% for the flat net income profile. Still, with rising income profiles, dynamic contracts without insurance of pre-age-25 health risk leave consumers facing a high share of the reclassification risk.

# 9 Comparison to Pauly, Kunreuther, and Hirth (1995) and Cochrane (1995)

Pauly, Kunreuther, and Hirth (1995) (PKH) propose guaranteed renewable contracts as a solution to prevent reclassification risk.[43] In contrast to the optimal long-term contracts described in Proposition 1, PKH aim to design a policy that provides full insurance in each period and guarantees that the consumer can renew in the future at the same premium as would be offered to the healthiest consumer type at that age. The idea is that a consumer with such a policy never wishes to lapse and faces no uncertainty in their consumption (i.e,

[43] Pauly, Kunreuther, and Hirth (1995) refer to their policies as guaranteed renewable contracts, but (as they note) effectively treat them as guaranteed premium path contracts. Actual "guaranteed renewable" contracts merely state that the consumer has a right to renew at a rate at the insurer's discretion, but that must be the same as what the insurer offers to all other consumers in the same policy. This discretion potentially allows the insurer to induce lapsation by raising premiums or degrading quality (effectively cancelling the policy) and to then re-enroll consumers in new policies at rates based on their health states. In practice, in the pre-ACA world, an insurer's ability to do this varied across states due to differences in state regulatory stringency.

no reclassification risk). In the context of our model, these policies provide a consumer who starts at age 25 in the healthiest possible state ($\Lambda_{25} = (1,1)$) with the guaranteed consumption path $\{y_t - p_t\}_{t=1}^{T}$, where the period $t$ premium $p_t$ is (we denote by $\Lambda_t = (1,1)$ the healthiest possible state in period $t$):[44]

$$p_t = \mathbb{E}[m_t|\Lambda_t = (1,1)] + \sum_{\tau>t} \delta^{\tau-t}\{\mathbb{E}[m_\tau|\Lambda_t = (1,1)] - \mathbb{E}[m_\tau|\Lambda_{t+1} = (1,1)]\} \text{ for } t = 1, ..., T \quad (8)$$

To understand this formula, consider a case in which $T = 2$. In period 2 (the last period), the consumer pays a premium equal to $p_2 = \mathbb{E}[m_2|\Lambda_2 = (1,1)]$, the expected medical expenses of the healthiest possible period-2 consumer. In period 1, the consumer pays a premium of

$$\begin{aligned} p_1 &= \mathbb{E}[m_1|\Lambda_1 = (1,1)] + \delta\{\mathbb{E}[m_2|\Lambda_1 = (1,1)] - \mathbb{E}[m_2|\Lambda_2 = (1,1)]\} \\ &= \mathbb{E}[m_1|\Lambda_1 = (1,1)] + \delta \Pr(\Lambda_2 \neq (1,1)|\Lambda_1 = (1,1))\{\mathbb{E}[m_2|\Lambda_2 \neq (1,1)] - \mathbb{E}[m_2|\Lambda_2 = (1,1)]\} \end{aligned}$$

The first term is the consumer's expected period-1 medical costs (since he starts with $\Lambda_1 = (1,1)$), while the second term is the prepayment of the expected period-2 discount being offered to the consumer (which he enjoys when it turns out that $\Lambda_2 \neq (1,1)$). In a problem in which $T = 3$, period 2 and 3 premiums would be as just described, while in the first period, in addition to a premium covering current-period expected medical expenses, the consumer would prepay the value of the expected period 2 premium discount,

$$\sum_{\tau=2,3} \delta^{\tau-1}\{\mathbb{E}[m_\tau|\Lambda_1 = (1,1)] - \mathbb{E}[m_\tau|\Lambda_2 = (1,1)]\}.$$

The fact that the PKH policies differ from those described in Proposition 1 points to two limitations of the PKH approach. First, unlike the contracts described in Proposition 1, the PKH contracts do not optimally balance the benefits of reducing reclassification risk against the costs of front-loading; indeed, as formula (8) makes clear, the PKH contract is unaffected by a consumer's income profile and risk preferences. The PKH contracts go to the extreme of completely preventing reclassification risk, resulting in a fully deterministic consumption profile but excessively low initial consumption.[45] Second, because PKH contracts do not optimize in this way, they are actually not lapsation-proof in periods $t \in [2, T-1]$ if insurers

[44] We provide the derivation in Appendix C. PKH focus on the case in which the consumer starts in the healthiest possible state.

[45] PKH's Proposition 2 asserts that their contract is Pareto optimal, but their argument assumes that "capital markets are perfect," when the motivation for considering these contracts to begin with was precisely the presence of capital market imperfections that prevent the payment of all lifetime premiums up front [see Pauly, Kunreuther, and Hirth (1995, p. 146)].

offer optimal dynamic contracts: the healthiest types would actually be drawn away by a firm offering the contracts described in Proposition 1.[46]

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

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

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.

One possible solution to this problem would be to have the consumer pre-pay the maximum possible ex post payment up front as part of his premium.[47] In each period $t$ the consumer would pay a premium of

$$\mathbb{E}[m_t|\Lambda_t] + \sum_{\tau>t} \delta^{\tau-t}\{\mathbb{E}[m_\tau|\Lambda_t] - \mathbb{E}[m_\tau|\Lambda_{t+1} = (1,1)]\} \tag{9}$$

and in each period $t+1$ (for $t < T$) the insurer would pay the consumer the non-negative amount

$$\begin{aligned}
\text{Payment} &= \sum_{\tau>t} \delta^{\tau-(t+1)}\{\mathbb{E}[m_\tau|\Lambda_{t+1}] - \mathbb{E}[m_\tau|\Lambda_t]\} + \sum_{\tau>t} \delta^{\tau-(t+1)}\{\mathbb{E}[m_\tau|\Lambda_t] - \mathbb{E}[m_\tau|\Lambda_{t+1} = (1,1)]\} \\
&= \sum_{\tau>t} \delta^{\tau-(t+1)}\{\mathbb{E}[m_\tau|\Lambda_{t+1}] - \mathbb{E}[m_\tau|\Lambda_{t+1} = (1,1)]\} \geq 0, \qquad (10)
\end{aligned}$$

equal to the change in expected medical expenses plus the repayment (with interest) of the second term in (9). Subtracting the period $t$ payment [given by expression (10) modified to be for period $t$ rather than $t+1$] from the period $t$ premium (9), 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 (8).

[46] This lapsation problem in the PKH contract arises only if $T \geq 3$.

[47] Cochrane (1995) proposes a different approach: an account that has to be used to receive and make these premium insurance payments. Unfortunately, Cochrane's assertion that such an account would never run into a negative balance for a consumer who starts out healthy is incorrect. For example, 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.

### 9.1 Empirical Comparison of PKH and Optimal Dynamic Contracts

Using formula (8), 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.[48] 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: 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.[49]

## 10 Conclusion

Long-term health insurance contracts offer a way to reduce reclassification risk without requiring community rating as in the ACA. Our results show that the value of a system that fully implements optimal dynamic contracts for men in Utah depends crucially on (i) whether there is government insurance of pre-age-25 health risk and (ii) the steepness of consumer income profiles. A lack of pre-age-25 risk insurance greatly reduces the appeal of dynamic contracts, while, whether or not such insurance is in place, the appeal of these contracts is greater if lifetime income profiles are flatter. With pre-age-25 health risk insurance in place, consumers with flat net income prefer dynamic contracts to the ACA-like environment we study, but consumers with steeper income profiles prefer the ACA-like environment. When

[48] At age 25 the value of the second term in equation (8), representing the premium pre-payment that is required in the PKH contract, is \$1,530. This amount divided by $\delta$ $(= 0.975)$ 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).

[49] 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.

we allow for meaningful switching cost (as empirical work has shown are relevant in practice) or lower risk aversion levels dynamic contracts become more attractive.

These results illustrate that there are certain plausible scenarios where dynamic contracts could improve welfare relative to an ACA-like environment. However, in practice, unlike in auto insurance or life insurance, such contracts have been very rare in health insurance markets. There are some potential practical impediments that are outside the scope of our model that could limit the viability of such contracts, which we now discuss.

One concern is that firms may have difficulty forecasting future medical cost levels, an issue that does not arise in markets such as life insurance in which long-term contracts are prevalent. This risk is not fully diversifiable.[50] This issue could be solved (alleviated), by indexing future guaranteed premiums to medical cost inflation indices in a granular manner, something that, e.g., is currently done in the German private health insurance market.

Another potential problem is that consumer lock-in might lead to quality degradation by insurers. This is something that was a major concern in the pre-ACA individual market for insured consumers with pre-existing conditions. While there are a number of ways to regulate product quality (on financial and non-financial dimensions) this is a concern that is potentially difficult to fully resolve. Cutting against this concern is the possibility that insurers' quality incentives would actually be enhanced on some dimensions, as they would have increased incentives to promote long-term health. Lock-in could also reduce a consumer's ability to re-match with firms if firm-specific preferences change (Atal (2015)). This problem would be greatly reduced if health insurance products were purely financial.

The most serious limitation on the use and benefits of dynamic contracts in the U.S. is the short durations of insurance need in the individual market. Given the current tax-advantage for employer-based insurance, consumers may arrive only when old, or in between jobs. For example, in the pre-ACA world, while some consumers purchased individual insurance over long periods of time, many others used it as a short-term solution between employment spells, leading median duration in the individual market in one study to be less than two years [Marquis et al (2006); see also Herring, Song, and Pauly (2008)].[51] The same is currently true of the ACA individual market exchanges. Short durations greatly reduce the

[50] The need to forecast could also introduce "winner's curse" type concerns, as firms who attract a lot of business would tend to be those whose forecasts of future medical cost inflation are unreasonably low.

[51] These short durations may well explain the apparent absence of explicitly dynamic contracts in the pre-ACA world. However, it is worth noting that, prior to the ACA, in most states insurers faced guaranteed renewability regulations that prevented them from re-pricing a policy to continuing customers on an individual basis [Patel and Pauly (2002)], and that such regulations did limit the reclassification risk that consumers faced once enrolled in a policy [Marquis and Buntin (2006), Herring and Pauly (2006), Herring, Song, and Pauly (2008)]. Fleitas, Gowrisankaran, and Losasso (2018) document a similar fact for the small group insurance market.

benefits of a long-term contract.[52] In addition, those older consumers newly arriving to the individual market with pre-existing conditions (perhaps because of a job loss) would still face reclassification risk, much as in our discussion of unhealthy 25 year-old consumers in Section 6.1, perhaps necessitating some sort of government insurance (such as high-risk pool subsidies). Removing the employer tax exemption for health insurance is one oft-discussed policy that would help promote the robustness of the individual market, whether in the ACA exchanges or in an individual market for dynamic contracts like we consider here.

In summary, our analysis shows scenarios under which long-term dynamic contracts may be welfare improving relative to a range of alternatives. In practice, several complementary regulations are likely important to help such contracts flourish. One key factor in our analysis that helps dynamic contracts, which was not present pre-ACA, is government insurance of pre-age-25 health risk. Such insurance is crucial to prevent consumers from facing significant pre-age-25 reclassification risk. Outside of our model, it is clear that, as for the ACA exchanges, removing the employer tax exemption will improve robustness of the individual market and meaningfully increase the length of consumer spells in that market. As this suggests, extending our analysis to allow for multiple market layers (e.g. employer markets, Medicare, Medicaid) that exist alongside an individual market with dynamic contracts is an important avenue for future work.

# References


[1] Abramitzky, R.(2008) "The Limits of Equality: Evidence from the Israeli Kibbutz," *Quarterly Journal of Economics* 123: 1111-59.

[2] Akerlof, G. (1970) "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics* 84: 488-500.

[3] Atal, J. P. (2016), "Lock-In in Dynamic Health Insurance Contracts: Evidence from Chile," working paper.

[4] Atal, J.P, H. Fang, M. Karlsson, and N. Ziebarth (2019), "Long-term Health Insurance: Theory Meets Evidence," working paper.

[5] Barro, R. (2006), "Rare Disasters, Asset Prices, and Welfare Costs," *Quarterly Journal of Economics* 121: 832-66.


[52] Introducing an exogenous probability of break up into our model is equivalent to lowering the discount rate, provided that separation payments upon break up cannot be made. With a discount factor of 0.5 (which generates an expected duration of 2 years), optimal dynamic contracts close only 8.4% of the gap between spot contracts and the no-borrowing/no-savings benchmark for a consumer with flat net income.

[6] Brown, J., M. Duggan, I. Kuzioemko, and W. Woolsten, "How Does Risk Selection Respond to Risk Adjustment? Evidence from the Medicare Advantage Program," *American Economic Review* 104: 3335-64.

[7] Browne, M. and A. Hoffmann. (2013) "One-Sided Commitment in Dynamic Insurance Contracts: Evidence from Private Health Insurance in Germany" *Journal of Risk and Uncertainty* 46: 81-112.

[8] Bundorf, K., J. Levin, and N. Mahoney (2012), "Pricing and Welfare in Health Plan Choice," *American Economic Review* 102: 3214-48.

[9] Cardon, J. and I. Hendel (2001), "Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey," *RAND Journal of Economics* 32: 408-27.

[10] Carlin, C. and R. Town (2009). "Adverse Selection, Welfare and Optimal Pricing of Employer Sponsored Health Plans," working paper.

[11] Chiappori, P. and B. Salanie (2000), "Testing for Asymmetric Information in Insurance Markets," *Journal of Political Economy* 108: 56-78.

[12] Cochrane, J. (1995), "Time-Consistent Health Insurance," *Journal of Political Economy* 103: 445-73.

[13] Collier, B, H. Kunreuther, E.O. Michel-Kerjan, and D. Schwartz (2017), "Risk Preferences in Small and Large Stakes: Evidence from Insurance Contract Decisions," working paper.

[14] DeGaridel-Thoron, T. (2005), "Welfare-Improving Asymmetric Information in Dynamic Insurance Markets," *Journal of Political Economy* 113: 121-50.

[15] Farhi, E. and I.Werning (2013), "Insurance and TAxation over the Life Cycle," *Review of Economic Studies* 80: 596-635.

[16] Finkelstein, A, K. McGarry and A. Sufi. (2005) "Dynamic Inefficiencies in Insurance Markets: Evidence from Long-Term Care Insurance," *The American Economic Review* 95: 224-8.

[17] Fudenberg, D., B. Holmstrom, and P. Milgrom (1990), "Short-term Contracts and Long-term Agency Relationships," *Journal of Economic Theory* 51: 1-31.

[18] Golosov, Troshkin, and Tsyvinski (2016), "Redistribution and Social Insurance," *American Economic Review* 106: 359-86.

[19] Handel, B. (2013), "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts," *American Economic Review* 103: 2643-82.

[20] Hendren, N. (2013), "Private Information and Insurance Rejections," *Econometrica* 81: 1713-62.

[21] Herring, B. and M.V. Pauly (2006), "Incentive-Compatible Guaranteed Renewable Health Insurance Premiums," *Journal of Health Economics* 25: 395-417.

[22] Ho, K., J. Hogan, and F. Scott Morton (2016), "The Impact of Consumer Inattention on Insurer Pricing in the Medicare Part D Program," NBER working paper no. 21028.

[23] Fleitas, S., G. Gowrisankaran, and A.L. Sasso (2018), "Reclassification Risk in the Small Group Health Insurance Market," working paper.

[24] Glazer, J., T. McGuire, and J. Shi (2014), "Risk Adjustment of Health Payment Plans to Correct Inefficient Plan Choice From Adverse Selection" NBER working paper no. 19998.

[25] Handel, B. (2013), "Adverse Selection and Switching Costs in Health Insurance Markets: When Nudging Hurts," *American Economic Review* 103: 2643-82.

[26] Handel, B., I. Hendel and M. Whinston (2015), "Equilibria in Health Exchanges: Adverse Selection vs. Reclassification Risk," *Econometrica* 83: 1261-1313.

[27] Harris, M. and B. Holmstrom (1982), "A Theory of Wage Dynamics," *Review of Economic Studies,* 40: 315-33.

[28] Hendel, I. (2016), "Dynamic Selection and Reclassification Risk: Theory and Empirics," working paper.

[29] Hendel, I. and A. Lizzeri (2003), "The Role of Commitment in Dynamic Contracts: Evidence from Life Insurance," *Quarterly Journal of Economics* 118: 299-327.

[30] Herring, B. and M. Pauly (2006), "Incentive-Compatible Guaranteed Renewable Health Insurance Premiums," *Journal of Health Economics* 25: 395-417.

[31] Herring, B., X. Song, and M. Pauly (2008), "Changes in Coverage in the Individual and Group Health Insurance Markets and the Effect of Health Status," U.S. Department of Health and Human Services Report, Office of the Assistant Secretary for Planning and Evaluation.

[32] Hofmann, A. and M. Browne (2013), "One-Sided Commitment in Dynamic Insurance Contracts: Evidence from Private Health Insurance in Germany," *Journal of Risk and Uncertainty* 46: 81-112.

[33] Illanes, G. (2017), "Switching Costs in Pension Plan Choice," working paper.

[34] Koch, T. (2011), "One Pool to Insure Them All? Age, Risk, and the Price(s) of Medical Insurance," working paper.

[35] Lavetti, K., DeLeire, T. and Ziebarth, N. (2018), "How Do Low-Income Enrollees in The Affordable Care Act Exchanges Respond to Cost-Sharing?," working paper.

[36] Marquis, M.S. and M.B. Buntin (2006), "How Much Risk Pooling Is There in the Individual Insurance Market?," *Health Services Research* 41: 1782-1800.

[37] Marquis, M.S., M.B. Buntin, J.J. Escarce, K. Kapur, T.A. Louis, and J.M. Yegian (2006), "Consumer Decision Making in the Individual Health Insurance Market," *Health Affairs* 25: w226-34.

[38] Pauly, M., H. Kunreuther and R. Hirth (1995), "Guaranteed Renewability in Insurance," *Journal of Risk and Uncertainty* 10: 143-56.

[39] Pauly, M., A. Percy, and B. Herring (1999), "Individual vs. Job-Based Health Insurance: Weighing the Pros and Cons," *Health Affairs* 18: 28-44.

[40] Rabin, M. (2000), "Risk Aversion and Expected-Utility Theory: A Calibration Theorem," *Econometrica* 68: 1281-92.

[41] Rothschild, M. and J. E. Stiglitz (1976), "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *Quarterly Journal of Economics* 90: 629-49.

# 11 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 $\lambda_t$, 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 $t$, $\Lambda_t^1$, which we refer to as the consumer's period-$t$ "health state," and is given by $f(\Lambda_{t'}^t | \Lambda_t^1)$.[53] Finally, we denote by $\left\langle \Lambda_{t'}^t, \Lambda_{t''}^{t'+1} \right\rangle$ a health history constructed by putting together $\Lambda_{t'}^t$ and $\Lambda_{t''}^{t'+1}$.

## 11.1 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 $\Lambda_t^1$, an optimal contract will depend only this, and not on previous medical expense realizations. We therefore denote a contract signed with the consumer at health history $\Lambda_t^1$ by $c_{\Lambda_t^1}(\cdot)$.[54] 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 $(\Lambda_{t'+1}^{t+1}, M_{t'}^t)$.[55] 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

$$\Sigma_{\tau=t}^T \delta^{\tau-t} \left( [y_\tau - \mathbb{E}[m_\tau | \Lambda_t^1] - \mathbb{E}[c_{\Lambda_t^1}(\Lambda_{\tau+1}^{t+1}, M_\tau^t) | \Lambda_t^1] \right) = -S \qquad (11)$$

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 $\Lambda_t^1$ that break even with subsidy $S$ by $B^S(\Lambda_t^1)$.

[53] 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.

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

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

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

$$V_{\Lambda_t^1}(c_{\Lambda_t^1}(\cdot)) = \Sigma_{\tau=t}^{T}\delta^{\tau-t}[\mathbb{E}[u(c_{\Lambda_t^1}(\Lambda_{\tau+1}^{t+1}, M_\tau^t))|\Lambda_t^1] \tag{12}$$

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

$$\max_{c_{\Lambda_t^1}(\cdot)\in B^{S_t}(\Lambda_t^1)} V_{\Lambda_t^1}(c_{\Lambda_t^1}(\cdot)) \tag{13}$$

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

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

Note that equation (13) provides a recursive definition of the optimal contract. The constraint in this definition makes sure that at no continuation health history $\Lambda_{t'}^{t+1}$ does the customer prefer to lapse to $c^*_{\langle\Lambda_t^1,\Lambda_{t'}^{t+1}\rangle}(\cdot|-\sigma)$, the optimal contract starting at health state $\langle\Lambda_t^1, \Lambda_{t'}^{t+1}\rangle$ 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 $c^*_{\Lambda_t^1}(\cdot|S)$ must effectively compensate the consumer for the fact that he incurs the switching cost $\sigma$.[56] 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 $c_{(\Lambda_t^1,\Lambda_{t'}^{t+1})}(\cdot)$ that would at least break even given the need to compensate the consumer for his switching cost, and that also satisfies no-lapsation.

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

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 1** *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 1.** 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 $\Lambda^1_t$, 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 consider contract $c_{\Lambda^1_t}(\cdot)$ 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

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

and otherwise,

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

Given that the utility function $u(\cdot)$ is strictly concave, this consumption-smoothing modification will imply that the customer strictly prefers $c_{\Lambda^1_t}(\cdot)$ over $c^*_{\Lambda^1_t}(\cdot|S)$. 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 $\Lambda^1_t$. Thus, contract $c_{\Lambda^1_t}(\cdot)$ also satisfies no-lapsation. But this contradicts the optimality of $c^*_{\Lambda^1_t}(\cdot|S)$.$\blacksquare$

Given lemma (1), we can simplify notation and write contracts in the form of $c_{\Lambda^1_t}(\Lambda^{t+1}_{t'}|S)$. 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 $\Lambda^1_{t'} = \left\langle \Lambda^1_t, \Lambda^{t+1}_{t'} \right\rangle$ that has been reached at date $t' \geq t$; hence in the form of $c_{\Lambda^1_t}(\Lambda^1_{t'}|S)$.

We introduce two more notations on comparing contracts to one another before we turn to the proposition and its proof. First, for two contracts $c_{\Lambda^1_t}(\cdot)$ and $\hat{c}_{\Lambda^1_t}(\cdot)$ offered at the same health state $\Lambda_t^1$, 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 $\Lambda_t^1$, 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 $\Lambda_{t'}^1 = \langle \Lambda_t^1, \Lambda_{t'}^{t+1} \rangle$ 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 $c_{\Lambda_t^1|\Lambda_{t'}^{t+1}}(\cdot) \geq \hat{c}_{\Lambda_t^1|\Lambda_{t'}^{t+1}}(\cdot)$.[57]

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

## 11.2 Proposition and Proof

We establish the following result, from which Proposition 1 follows as the special case where the switching cost $\sigma$ equals zero.[58]

**Proposition 3** *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 the following certain consumption level:*

$$c^*_{\Lambda_t^1}(\Lambda_{t'}^1) = \max\{c^*_{\Lambda_t^1}(\Lambda_t^1), \max_{\tau \in \{t+1,\ldots,t'\}} c^*_{\langle \Lambda_t^1, \Lambda_\tau^{t+1} \rangle}(\langle \Lambda_t^1, \Lambda_\tau^{t+1} \rangle)| - \sigma)\}. \tag{14}$$

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

The proof strategy is based on strong induction: We assume the proposition is true for the optimal contracts $c^*_{\Lambda_{t'}^1}(\cdot)$ at all $\Lambda_{t'}^1$ with $t' > t$, and then show it is also true for the period-$t$ optimal contracts $c^*_{\Lambda_t^1}(\cdot)$ for any $\Lambda_t^1$. To establish the result, we show that if for some $\Lambda_t^1$, optimal contract $c^*_{\Lambda_t^1}(\cdot)$ does not satisfy (14), then there is a "modification of" $c^*_{\Lambda_t^1}(\cdot)$ that (i) is strictly preferred to $c^*_{\Lambda_t^1}(\cdot)$ 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 3** *Let $\min\{t', t''\} \geq t$. We say contract $\hat{c}_{\Lambda_t^1}(\cdot)$ is an $\varepsilon$-transfer, from $\Lambda_{t'}^1$ to $\Lambda_{t''}^1$, on contract $c_{\Lambda_t^1}(\cdot)$, and write $\hat{c}_{\Lambda_t^1}(\cdot) = tr[c_{\Lambda_t^1}(\cdot), \varepsilon, \Lambda_{t'}^1, \Lambda_{t''}^1]$ if:*

1. $\hat{c}_{\Lambda_t^1}(\Lambda_{t'}^1) = c_{\Lambda_t^1}(\Lambda_{t'}^1) - \varepsilon$

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

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

2. $\hat{c}_{\Lambda_t^1}(\Lambda_{t''}^1) = c_{\Lambda_t^1}(\Lambda_{t''}^1) + \left[\varepsilon \times \frac{f(\Lambda_{t'}^{t+1}|\Lambda_t^1)}{f(\Lambda_{t''}^{t+1}|\Lambda_t^1)} \times \delta^{t'-t''}\right]$

3. *For all* $\tau \geq t$ *and* $\Lambda_\tau^t \notin \{\Lambda_{t'}^t, \Lambda_{t''}^t\}$, *we have* $\hat{c}_{\Lambda_t^1}(\Lambda_\tau^t) = c_{\Lambda_t^1}(\Lambda_\tau^t)$

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

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

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

Remark 1 follows immediately from the fact that the $\varepsilon$-transfer does not change the expected discounted consumption in the contract, while Remark 2 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* $c^*_{\Lambda_t^1}(\cdot|S_t)$ *after the realization of health history* $\Lambda_{t'}^1 = \langle \Lambda_t^1, \Lambda_{t'}^{t+1}\rangle$. *Formally:*

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

*Then, the following is true:*

$$c^*_{\Lambda_t^1|\Lambda_{t'}^{t+1}}(\cdot|S_t) = c^*_{\langle \Lambda_t^1, \Lambda_{t'}^{t+1}\rangle}(\cdot|S_{t'}), \tag{15}$$

*where*

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

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

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 2** *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*

$$c^*_{\Lambda^1_t}(\left\langle\Lambda^1_t, \Lambda^{t+1}_{t'}\right\rangle|S_t) = \max\{\bar{c}, c^*_{\left\langle\Lambda^1_t,\lambda_{t+1}\right\rangle}(\left\langle\Lambda^1_t, \Lambda^{t+1}_{t'}\right\rangle| - \sigma)\}. \tag{16}$$

In words, Lemma 2 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 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 (16) of the lemma implies that any two optimal contracts signed at time $t$ and health history $\Lambda^1_t$, 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 2.** The proof goes by induction. For $t = T$, the result is straightforward, given that there is no period after $t = T$; at that point, $c_{\Lambda^1_T}(\Lambda^1_T|S) = y_T - \mathbb{E}[m_T|\lambda_T] + S$. We now turn to the proof for $t < T$, assuming, by way of induction, that the result holds for any $\tau > t$ and any $S_\tau$. We begin by showing that (16) holds for any period $t+1$ history $\langle\Lambda^1_t, \lambda_{t+1}\rangle$; i.e., that

$$c^*_{\Lambda^1_t|\lambda_{t+1}}(\left\langle\Lambda^1_t, \lambda_{t+1}\right\rangle|S_t) = \max\{c^*_{\Lambda^1_t}(\Lambda^1_t|S_t), c^*_{\left\langle\Lambda^1_t,\lambda_{t+1}\right\rangle}(\left\langle\Lambda^1_t, \lambda_{t+1}\right\rangle| - \sigma)\}. \tag{17}$$

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

<u>Case 1.</u> *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:

$$V_{\Lambda^1_t}(c^*_{\Lambda^1_t|\lambda_{t+1}}(\cdot|S_t)) = V_{\Lambda^1_t}(c^*_{\left\langle\Lambda^1_t,\lambda_{t+1}\right\rangle}(\cdot| - \sigma)). \tag{18}$$

Note that by Claim 1, the continuation contract $c^*_{\Lambda^1_t|\Lambda^{t+1}_{t'}}(\cdot|S_t)$ is itself the optimal contract $c^*_{\left\langle\Lambda^1_t,\lambda_{t+1}\right\rangle}(\cdot|S_{t+1})$ for some $S_{t+1}$. Thus, (18) implies that

$$c^*_{\Lambda^1_t|\lambda_{t+1}}(\cdot|S_t) = c^*_{\left\langle\Lambda^1_t,\lambda_{t+1}\right\rangle}(\cdot| - \sigma). \tag{19}$$

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

$$c^*_{\Lambda^1_t}(\Lambda^1_t|S_t) \leq c^*_{\Lambda^1_t}(\langle \Lambda^1_t, \lambda_{t+1}\rangle |S_t) = c^*_{\langle \Lambda^1_t, \lambda_{t+1}\rangle}(\langle \Lambda^1_t, \lambda_{t+1}\rangle | -\sigma), \tag{20}$$

otherwise an $\varepsilon$-transfer from the immediate consumption at history $\Lambda^1_t$ 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, (17) 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,

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

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

$$c^*_{\Lambda^1_t|\lambda_{t+1}}(\langle \Delta^1_t, \lambda_{t+1}, \Lambda^{t+2}_{t'}\rangle |S_t) = \max\{\bar{c}, c^*_{\langle \Lambda^1_t, \lambda_{t+1}\rangle}(\langle \Delta^1_t, \lambda_{t+1}, \Lambda^{t+2}_{t'}\rangle | -\sigma)\}. \tag{22}$$

But this, combined with inequality (21), tells us it must be that

$$\bar{c} > c^*_{\langle \Lambda^1_t, \lambda_{t+1}\rangle}(\langle \Lambda^1_t, \lambda_{t+1}\rangle | -\sigma). \tag{23}$$

We now claim that

$$\bar{c} = c^*_{\Lambda^1_t}(\Lambda^1_t|S_t). \tag{24}$$

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 $\varepsilon$-transfer from the history $(\Lambda^1_t, \lambda_{t+1})$ to the immediate history $\Lambda^1_t$ will increase the consumer's expected utility, will not change the expected profit from the contracts, and will preserve no-lapsation if $\varepsilon$ is small enough, given (21). 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 $\varepsilon$-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 $\Lambda^1_t$. Thus, (17) 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 $\lambda_{t+1}$, no matter whether no-lapsation is binding or not, (17) holds. Next, we combine (17) 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 $\left\langle \Lambda_t^1, \Lambda_{t'}^{t+1} \right\rangle$ with $t' > t+1$. By Claim 1, we know that for some appropriate $S_{t+1}$, we have:

$$c^*_{\Lambda_t^1}(\left\langle \Lambda_t^1, \lambda_{t+1} \right\rangle |S_t) = c^*_{\left\langle \Lambda_t^1, \lambda_{t+1} \right\rangle}(\left\langle \Lambda_t^1, \lambda_{t+1} \right\rangle |S_{t+1}) \tag{25}$$

and

$$c^*_{\Lambda_t^1}(\left\langle \Lambda_t^1, \lambda_{t+1}, \Lambda_{t'}^{t+2} \right\rangle |S_t) = c^*_{\left\langle \Lambda_t^1, \lambda_{t+1} \right\rangle}(\left\langle \Lambda_t^1, \lambda_{t+1}, \Lambda_{t'}^{t+2} \right\rangle |S_{t+1}) \tag{26}$$

By induction, we know that

$$c^*_{(\Lambda_t^1, \lambda_{t+1})}(\left\langle \Lambda_t^1, \lambda_{t+1}, \Lambda_{t'}^{t+2} \right\rangle |S_{t+1}) = \max\{c^*_{\left\langle \Lambda_t^1, \lambda_{t+1} \right\rangle}(\left\langle \Lambda_t^1, \lambda_{t+1} \right\rangle |S_{t+1}), c^*_{\left\langle \Lambda_t^1, \lambda_{t+1}, \lambda_{t+2} \right\rangle}(\left\langle \Lambda_t^1, \lambda_{t+1}, \Lambda_{t'}^{t+2} \right\rangle |-\sigma)\} \tag{27}$$

Replacing into equation (27) from (25) and (26), we get:

$$c^*_{\Lambda_t^1}(\left\langle \Lambda_t^1, \lambda_{t+1}, \Lambda_{t'}^{t+2} \right\rangle |S_t) = \max\{c^*_{\Lambda_t^1}(\left\langle \Lambda_t^1, \lambda_{t+1} \right\rangle |S_t), c^*_{\left\langle \Lambda_t^1, \lambda_{t+1}, \lambda_{t+2} \right\rangle}(\left\langle \Lambda_t^1, \lambda_{t+1}, \Lambda_{t'}^{t+2} \right\rangle | - \sigma)\} \tag{28}$$

Now, substituting for $c^*_{\Lambda_t^1}(\langle \Lambda_t^1, \lambda_{t+1} \rangle |S_t)$ from (17), we get:

$$\begin{aligned} & c^*_{\Lambda_t^1}(\left\langle \Lambda_t^1, \lambda_{t+1}, \Lambda_{t'}^{t+2} \right\rangle |S_t) \qquad (29) \\ = \; & \max\{\max\{c^*_{\Lambda_t^1}(\Lambda_t^1|S_t), c^*_{(\Lambda_t^1, \lambda_{t+1})}(\left\langle \Lambda_t^1, \lambda_{t+1} \right\rangle | - \sigma)\}, c^*_{\left\langle \Lambda_t^1, \lambda_{t+1}, \lambda_{t+2} \right\rangle}(\left\langle \Lambda_t^1, \lambda_{t+1}, \Lambda_{t'}^{t+2} \right\rangle | - \sigma)\} \\ = \; & \max\{c^*_{\Lambda_t^1}(\Lambda_t^1|S_t), \max\{c^*_{(\Lambda_t^1, \lambda_{t+1})}(\left\langle \Lambda_t^1, \lambda_{t+1} \right\rangle | - \sigma), c^*_{\left\langle \Lambda_t^1, \lambda_{t+1}, \lambda_{t+2} \right\rangle}(\left\langle \Lambda_t^1, \lambda_{t+1}, \Lambda_{t'}^{t+2} \right\rangle | - \sigma)\}\} \end{aligned}$$

But by our induction assumption, the inner maximum equals $c^*_{\left\langle \Lambda_t^1, \lambda_{t+1} \right\rangle}(\left\langle \Lambda_t^1, \lambda_{t+1}, \Lambda_{t'}^{t+2} \right\rangle | - \sigma)$; substituting this into (29) tells us that (16) holds for contracts signed in period $t$. Applying induction establishes the lemma.■

**Proof of Proposition 3.** Applying Lemma 2 to the special case of $S_t = 0$, we get that for any $\Lambda_t^1$ and $\Lambda_{t'}^{t+1}$ such that $f(\Lambda_{t'}^{t+1}|\Lambda_t^1) > 0$, we have that

$$c^*_{\Lambda_t^1}(\left\langle \Lambda_t^1, \Lambda_{t'}^{t+1} \right\rangle) = \max\{c^*_{\Lambda_t^1}(\Lambda_t^1), c^*_{\left\langle \Lambda_t^1, \lambda_{t+1} \right\rangle}(\left\langle \Lambda_t^1, \Lambda_{t'}^{t+1} \right\rangle | - \sigma)\}$$

Since Lemma 2 holds for $c^*_{\langle \Lambda_t^1, \lambda_{t+1} \rangle}(\cdot|-\sigma)$ as well, we can expand $c^*_{\langle \Lambda_t^1, \lambda_{t+1} \rangle}(\langle \Lambda_t^1, \Lambda_{t'}^{t+1} \rangle |-\sigma)$ 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 $f(\Lambda_{t'}^t|\Lambda_t^1) > 0$,

$$c^*_{\Lambda_t^1}(\Lambda_{t'}^1) = \max\{c^*_{\Lambda_t^1}(\Lambda_t^1), \max_{\tau \in \{t+1,...,t'\}} c^*_{\langle \Lambda_t^1, \Lambda_\tau^{t+1} \rangle}(\Lambda_\tau^1)| - \sigma)\}, \tag{30}$$

which is exactly the statement made in the proposition.■

# 12 Appendix B: Self-Selection

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

**Lemma 3** *Let $p = (p_\tau, ..., p_T)$ and $\widehat{p} = (\widehat{p}_\tau, ..., \widehat{p}_T)$ be guaranteed premium paths (that start in period $\tau$) such that $\widehat{p} \geq p$ and consider a type $\theta$ consumer who is in health state $\Lambda_\tau^1$ in period $\tau$. 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 $\Lambda_t^1$ 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 $\tau$ by the consumer.*

**Proof.** Let $U(t, \Lambda_t^1)$ and $\widehat{U}(t, \Lambda_t^1)$ denote the type $\theta$ consumer's continuation payoff in period $t$ at health state $\Lambda_t^1$ 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 $(t, \Lambda_t^1)$ given the consumer's optimal lapsation behavior under $p$ and $\widehat{p}$ respectively. Assumption (i) therefore says that $\Pi(\tau, \lambda_\tau) \geq 0$, 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 $(t, \Lambda_t^1)$: 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.[59] Next, the fact that $U(t, \Lambda_t^1) \geq \widehat{U}(t, \Lambda_t^1)$ for

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

all $(t, \Lambda_t^1)$ 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 $(\tau, \Lambda_\tau^1)$ 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 4** *Suppose that in each period $t \geq \tau$ the menu of contracts offered to a consumer who is in health state $\Lambda_t^1$ 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 $\theta$ consumer in health state $\Lambda_\tau^1$ will not secretly save when facing guaranteed premium path $p_\tau^{\theta*}(\Lambda_\tau^1)$.*

**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 $\Lambda_t^1$ 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 $\Delta(t, \Lambda_t^1) \in \mathbb{R}$.[60] 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 $\Lambda_t^1$ 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

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

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 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.[61] 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 5** *Suppose that in each period $t \geq \tau$ the menu of contracts offered to a consumer who is in health state $\Lambda_t^1$ 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 $\tau$ type $\theta'$ consumer in health state $\Lambda_\tau^1$ chooses the guaranteed premium path $p_\tau^{\theta *}(\Lambda_\tau^1)$ that is intended for a type $\theta$ consumer in health state $\Lambda_\tau^1$.*

**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 $\theta'$ starting in period $t$ is either (A) the same as for type $\theta$ (meaning, it is the same after any history of health states between periods $\tau$ 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 $\theta$ consumer.

Observe that the Induction Hypothesis holds if $t = T$, since then lapsation behavior is the same for type $\theta'$ as for type $\theta$ – both types lapse if and only if $E[m_T|\lambda_T] < p_T^\theta$, where $p_T^\theta$ 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

[61] Note that a consumer with utility function $u$ who arrives in period $t$ in health state $\Lambda_t^1$ 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 $\theta = (y^S, u)$.

which the consumer has not yet lapsed. Recall that by Lemma 4 the type $\theta$ consumer does not secretly save; however, the type $\theta'$ consumer may.

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

Suppose, instead, that state $\Lambda_{t-1}^1$ is one in which type $\theta$ would lapse in period $t-1$, choosing a contract with premium path $\widehat{p}$, while type $\theta'$ 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 $\theta'$. By the self-selection assumption, we know that the contract $\widehat{p}$ that type $\theta$ 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 $\theta$ 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 $\theta$'s income). Hence, by Lemma 3, if the type $\theta$ 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 when the type $\theta'$ 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 $\Lambda_t^1$ 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 $\theta'$ consumer than when facing a type $\theta$ consumer. So the Induction Hypothesis holds in period $t-1$, and hence – applying induction – in period $\tau+1$.

Finally, consider period $\tau$. The argument is similar to that above: If a type $\theta'$ consumer in health state $\Lambda_\tau^1$ chooses the premium path $p_\tau^{\theta*}(\Lambda_\tau^1)$ intended for a type $\theta$ consumer in health state $\Lambda_\tau^1$, the insurer's first period costs are the same as if a type $\theta$ consumer in health state $\Lambda_\tau^1$ 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 $\theta'$ consumer in health state $\Lambda_\tau^1$ chooses contract $p_\tau^{\theta*}(\Lambda_\tau^1)$ in period $\tau$. ■

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 $\theta$ 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 $\tau$; we argue that it is then also true for $\tau - 1$. Lemma 4 implies that a type $\theta$ 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 5 we know that if a type $\theta$ 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 $\theta'$ then insurers earn non-negative profits (all future insurers break even in both cases). But the policy intended for the type $\theta$ consumer maximizes the type $\theta$ 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 $\theta'$ was therefore feasible for type $\theta$, which implies that it cannot be preferred by type $\theta$.

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. ∎

# 13 Appendix C: PKH premiums

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 gurantees 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:

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

and for $t < T$,

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

The next result derives an expression for the premium paid in period $t$, corresponding to expression (8) in the text.

**Lemma 6** *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:

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

■