The Welfare Impact of Long-Term Health Insurance Contracts*

Ben Handel†, Igal Hendel‡, and Michael Whinston§

January 2017

Abstract

Reclassification risk is a major concern in health insurance. Regulation, like the ACA, prescribes community rating to contend with reclassification risk. However, community rating comes at the cost of adverse selection. We use a rich data set with individual-level information on health risk to empirically study an alternative solution: dynamic contracts. We compare equilibria under four contractual arrangements: (i) spot contracts, (ii) first-best long-run contracts with two-sided commitment, (iii) long-term contracts with one-sided commitment, and (iv) ACA-like markets with spot contracts and community rating. Empirically, dynamic contracts with one-sided commitment deliver a substantial welfare gain over spot contracts by attenuating reclassification risk. Dynamic contracts achieve close to the first-best for consumers with flat net income paths who are willing to front-load payments to facilitate long-run insurance. Consumers with steeper net income growth over their lifetimes prefer community rating over dynamic contracts.

1 Introduction

Consumers face substantial health risks over their lifetime. Much of this risk involves the development of conditions, such as diabetes, heart disease, and cancer, that lead to higher expected medical expenses over significant periods of time. This fact can expose

* All authors are grateful for support from NSF grant SES-1259770 and Whinston also thanks prior support from the NSF and the Toulouse Network for Information Technology. We thank Soheil Ghili for outstanding research assistance.

†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

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.

The state-by-state health insurance exchanges set up under the Affordable Care Act (ACA), and many similar markets worldwide, respond to this problem through pricing regulations that enforce community rating and guaranteed issuance, thereby prohibiting discrimination against consumers who have developed "pre-existing conditions." Unfortunately, while such bans can eliminate reclassification risk, requiring identical pricing for consumers with different ex ante risk levels can create adverse selection, leading to under-provision of insurance [Handel, Hendel and Whinston (2015), Patel and Pauly (2002)].

In this paper, we explore the extent to which long-term contracts, without pricing regulation, could offer a way to reduce reclassification risk without incurring welfare losses from adverse selection. Specifically, we characterize optimal long-term contracts theoretically, and then use data on the preferences, income paths, and health transitions of a population of employees at a large firm to assess the welfare achievable through long-term contracts, and compare it to other possible approaches, such as the annual contracts with community rating present in the ACA.

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 symmetrically learn those shocks over time. If only spot contracts are available, risk-averse consumers 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 can commit to long-term contracts ex ante, prior to information revelation, then the efficient (first-best) allocation, which involves full insurance, is possible. However, in practice, consumer commitment over a long time horizon is unlikely.[1] As a result, the empirically relevant contracting environment is likely one with one-sided commitment on the part of insurers, which we assume to be the case. We assume as well the presence of capital market imperfections preventing consumers from borrowing, due for example to a lack of collateral.[2] As shown in Harris and Holmstrom (1983) and Hendel and Lizzeri (2003), the competitive equilibrium in markets with one-sided commitment only partially insures reclassification risk.

[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 premia up front.

We show theoretically that optimal unilateral commitment contracts offer consumers a minimum guaranteed consumption level over time, as a function of their risk preferences and income paths. This minimal guarantee is bumped up to a new level within a contract when necessary to meet competitive offers from other insurers. The improved terms are necessary to ensure that consumers won't leave the long-term contract after receiving a positive shock. These consumption floors are the counterpart of the downwardly rigid wages in Harris and Holmstrom (1983), whose model we generalize and apply to the context of health insurance. As there, optimal contracts involve "front-loading" – here premium payments 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 – when expressed as guaranteed renewable premium paths over time – are self-selective, in the sense that consumers with different lifetime income paths prefer the contracts designed for them.

We apply this model empirically using an individual-level panel data set on the medical claims of the employees (and their dependents) of a large firm. The key data are diagnostic codes of each individual in the sample, which we combine with professional software designed to predict future medical expenditures to produce a measure of an individual's health status. We thus observe the medical expense risk that consumers face within the typical one-year span of health insurance contracts. (Notably, we know the same information insurers use at underwriting.) A key ingredient to study dynamic contracts is the stochastic process that determines the evolution of health. Because we have multiple observations of individuals' health status we are able to estimate a long-run health status transition process that captures how this risk evolves over time. We use risk preference estimates from Handel, Hendel, and Whinston (2015), identified based on the insurance choices in these data, to assess consumers' preferences for consumption smoothing.

We first compute optimal contracts with one-sided insurer commitment using our data, to examine the premiums and the extent of front-loading associated with these contracts. For consumers with flat net income over time ("net income" is income minus expected medical costs), a healthy consumer at age 25 pays a premium of \$2,750 despite expected costs of only \$1,131 in that year. The extent of front-loading is inverted U-shaped: with medium risks front-loading the most. This occurs because the extent of front-loading at any point in time depends on both the current health state and the implications of that current state for future states (measured through our estimated health status transition matrix). Good health draws can afford the front-loading better, while bad draws benefit more from the future promises front-loading buys. The extent of front-loading is substantial. Over the first ten years of these long-term contracts, the average consumer contributes \$21,209 extra

through front-loading, in order to fund long-term insurance against health shocks leading to reclassification risk.

We also examine the structure of these contracts as a function of consumers' expected income paths. In addition to the case of flat net income, we use the income paths of managers at the firm to represent steeply increasing net income, the income paths of non-managers to reflect rising but flatter income paths, and we also consider teh case of a "downscaled manager" whose lifetime income path is proportionately scaled down from that of a manager to generate the same present discounted value as that of a non-manager. Downscaled managers, for example, with steeper income growth, have equilibrium contracts with noticeably lower front-loading than for the case of flat net income: for example, a manager in perfect health at age 25 front-loads only \$547 in that year, over \$1,000 lower than the \$1,619 that a consumer with flat net income front-loads at that same age and health state. This feature holds in general: conditional on age and health state, managers front-load much less early in the contract, primarily because their marginal value of incremental income is much higher in those periods than later in life, when their incomes are much higher. Thus, their income growth over time limits their desire to insure future health risks with current income. Nonetheless, managers' contracts still exhibit substantial consumption guarantees that protect against the risks of their health status worsening over time. The primary difference relative to flatter income paths is that the consumption guarantees of managers are lower, due to the lower degree of front-loading.

We then consider the welfare effects of these long-term contracts. Our welfare analysis compares the lifetime certainty equivalent achievable by optimal dynamic contracts with one-sided commitment to that of following outcomes: (i) the first best (equivalent to long-term contracts with two-sided commitment), (ii) spot contracts, (iii) full within-period insurance but without intertemporal transfers (i.e., borrowing or saving), which eliminates reclassification risk but offers no consumption smoothing, and (iv) ACA-style exchanges with spot contracts and community-rating [as studied in Handel, Hendel, and Whinston (2015)]. We consider welfare outcomes for the four different income paths described above: (1) flat net income, (2) non-manager income, (3) manager income, and (4) "downscaled" managers.

We investigate first the welfare loss from a market with only year-to-year spot contracts (and no community rating). Spot contracts result in a welfare loss between 14% and 40% relative to the first best, depending on the income profile, with larger losses for consumers with steeper income paths. Part of these losses come from the inability of spot contracts to smooth consumption over time. Since our interest is insurance, not consumption smoothing, we also use as a benchmark contracts that offer full insurance with premiums at each age equal to the average medical expenses of that age group [outcome (iii) listed above]. These contracts offer full insurance, without allowing consumption smoothing over time. Relative

to this alternative benchmark, we find that spot contracts imply a 10% to 14% welfare loss, representing the loss from being unable to insure reclassification risk.

We then assess the performance of optimal dynamic contracts with one-sided commitment. We find that, for consumers with flat net income, these dynamic contracts achieve close to the first best, closing 87.7% of the welfare gap between spot contracts and two-sided commitment. However, for steeper income profiles, optimal dynamic contracts bridge much less of this gap. For example, for downscaled managers, optimal dynamic contracts with one-sided commitment bridge only 10.8% of this welfare gap. They perform somewhat better relative to the case of full insurance with no consumption smoothing, closing 38% of the welfare loss suffered by a downscaled manager due to spot contracts.

We compare as well optimal dynamic contracts with one-sided commitment to ACA-like market regulation with spot contracts and community rating, where price discrimination is prohibited and consumers are mandated to purchase one of several levels of coverage. In ACA-like markets, community-rating fully insures reclassification risk, but creates adverse selection [studied in Handel, Hendel, and Whinston (2015)] leading to a welfare loss between 2% and 6% relative to the first best, depending on the income path considered. This welfare loss occurs because consumers' within-period event insurance unravels to the minimum allowable coverage of 60% actuarial value in equilibrium. Interestingly, we find that whether the ACA-like market is preferred from a welfare perspective depends on the income paths we consider. For rising income paths (non-managers, managers, and downscaled managers) the ACA-like environment does better than dynamic contracts, but for consumers with flat net income paths, dynamic contracts are preferred. Intuitively, the ACA-like environment is better for individuals who find front-loading costly.

We also explore several robustness checks and extensions. First, we explore the role that risk aversion plays in our analysis, and specifically focus on the case in which risk aversion is one-fifth the magnitude as what we estimated in our empirical context. 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. Second, we consider the effects of switching costs, as estimated in Handel (2013). By improving consumers' commitment to their current insurer, these costs can improve the insurance of reclassification risk. We find that a switching cost in the range of $5,000 eliminates most of the reclassification risk, increasing the welfare dynamic contracts deliver to the range of full insurance without consumption smoothing and above what we find for the ACA. Third, we examine the welfare optimal dynamic contracts deliver to consumers who arrive at age 25 in different health states. We find that consumers who are in poor initial health are significantly disadvantaged relative to how they do in alternative market designs, a fact that suggests the desirability of providing some form of insurance for this

risk if contracting does not occur until age 25. Insuring this pre-age 25 health risk would create a significant improvement in the gain of otimal dynamic contracts over spot contracts. Noenhteless, even with such insurance, optimal dynamic contracts (absent switching costs) perform worse than the ACA for the rising income paths we consider.

We are not the first to consider the use of long-term contracts as a means of addressing reclassification risk in insurance markets. Long-term contracts have been explored in life insurance markets by Hendel and Lizzeri (2003) who document that long-term life-insurance contracts involve front-loaded premiums, as the theory of optimal contracts with one-sided commitment predicts. 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.[3]

In the context of health insurance, Pauly, Kunreuther, and Hirth (1995) and Cochrane (1995) provide theoretical discussions of schemes to address reclassification risk. 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 both of these ideas, and their relation to our optimal contracts, in Section 8. Closest to our approach here is Herring and Pauly (2006), who conduct an empirical calibration of the Pauly, Kunreuther and Hirth (1995) model, using data from the Medical Expenditure Panel Survey (MEPS). After deriving the guaranteed renewable contract they compare its time path of premiums to the average premiums paid by age in the MEPS data. Compared to our analysis, they place a much simpler structure on the evolution of health and possible health states, and don't derive optimal contracts that take account of consumption-smoothing concerns.

One-sided dynamic health insurance contracts are offered in several countries, including Germany and Chile. Browne and Hoffmann (2013) study the German private health insurance (PHI) market and demonstrate that (i) front-loading of premiums generates lock-in of consumers, (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.[4] Compared to all of this previous work on long-term health insurance contracts, our

[3] Abramitzky (2010) applies a similar model to understand the evolution and existence of the kibbutz as an institution.

[4] Bundorf, Levin, and Mahoney (2012) investigate the implications of reclassification risk in a large-employer context in a short-run environment. See Hendel (2016) for a survey of the literature on reclassification risk.

work is unique in using data to derive optimal contracts with one-sided commitment, and in assessing the welfare impacts of these contracts.

The rest of the paper proceeds as follows. Section 2 presents our model and its equilibrium implications. Section 3 presents the data we use to quantify the positive and normative implications of different market designs. Section 4 briefly discusses computation. Section 5 presents our main positive and normative results. Section 6 discusses extensions. Section 7 discusses the relation of our work to that of Pauly, Kunreuther, and Hirth (1995) [and, by extension, Herring and Pauly (2006)] and Cochrane (1995). Section 8 concludes.

2 Model of Dynamic Insurance

We consider a dynamic insurance problem T periods long, with periods indexed $t = 1, ...T$. In the 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.

A consumer enters each period t characterized by her health state $\lambda_t \in \Lambda$, with greater λ_t indicating sicker individuals. The consumer's initial health state is λ_1. For periods $t < T$, state λ_t determines both the distribution of period t medical expenses, m_t, and the transition probabilities to the period $t + 1$ health state λ_{t+1}, according to the joint density $f_t(m_t, \lambda_{t+1}|\lambda_t)$. We denote by $\overline{f}_t(\lambda_{t+1}|\lambda_t)$ the marginal conditional density of λ_{t+1}. As future health is not relevant for the period T insurance problem, we need only specify the density $f_T(m_T|\lambda_T)$.[5]

An individual's health state λ_t is observed by both the individual and all insurance firms, namely, there is symmetric information and symmetric learning.[6] We make the following assumptions concerning the stochastic health process, which we later verify empirically:

Assumption A1 $\mathbb{E}[m_t|\lambda_t]$ is strictly increasing in λ_t.

Assumption A2 If $\lambda'_t > \lambda_t$, then $\overline{f}_t(\lambda_{t+1}|\lambda'_t)$ first-order stochastically dominates $\overline{f}_t(\lambda_{t+1}|\lambda_t)$.

Assumption A1 says that the state λ_t captures the consumer's period t health: larger λ_t implies greater period t expected medical expenses. Assumption A2 says that being in worse health in period t implies that the consumer's health state in period $t + 1$ will be worse, in the sense of first-order stochastic dominance.

[5] Incorporating health-dependent income is straightforward. We can allow for a link between different health states and income by allowing for an additional expense associated with the state. Equilibrium contracts would be determined in the same way as we describe below.

[6] Our assumption that all insurers have access to the same information would apply, for example, in cases in which there are all-payers claims databases. 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. See, for example, DeGaridel-Thoron (2005).

We assume that the insurance market is perfectly competitive, with risk-neutral firms who discount future cash flows using the discount factor $\delta \in (0,1)$. Consumer preferences for consumption each period are described by $u(\cdot)$, the consumer's Bernoulli utility function. A consumer's long-run 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.[7] Throughout we assume that consumers are unable to borrow to fund premium payments or other expenses.

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 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 ex ante expectation of period t medical expenses, $\mathbb{E}[m_t|\lambda_1]$). That is, it involves the constant consumption level

$$C^* = \left(\frac{1-\delta}{1-\delta^T}\right) \sum_{t=1}^{T} \delta^{t-1}(y_t - \mathbb{E}[m_t|\lambda_1]). \tag{1}$$

As is well known, if consumers and insurance firms could both commit to a long-term contract at the start of period 1, the competitive outcome 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, 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 her period t health state λ_t. This results in the period t consumption level $y_t - \mathbb{E}[m_t|\lambda_t]$. Because the consumer's period t health state λ_t is ex ante uncertain (given λ_1), this outcome faces the consumer with risk from an ex ante perspective. The consumer's constant certainty equivalent of this uncertain consumption path is the constant consumption level CE_{SPOT} such that

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

[7] The model readily generalizes to stochastic income – possibly dependent on the consumer's health state, provided that $\mathbb{E}[y_t - m_t|\lambda_t]$ is strictly decreasing in λ_t.

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

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

$$u(C^*_{NBNS}) = \left(\frac{1-\delta}{1-\delta^T}\right) \sum_{t=1}^{T} \delta^{t-1} u(y_t - \mathbb{E}[m_t|\lambda_1])). \tag{3}$$

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 can view a long-term contract as specifying the consumer's period t consumption level c_t as a function of the consumer's publicly-observed health history up to period t, $h^t = [\lambda_1, (m_1, \lambda_2), ...(m_t, \lambda_{t+1})]$.[9] (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 history h^t must give her 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 yield the 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 h^t is it possible to offer the consumer a 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 define the state to be the pair (λ_t, S_t) where λ_t is the consumer's current health state, and S_t is the loss that the insurer is allowed to sustain going forward (i.e., S_t is the subsidy for

[9] 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.

future insurance). This is a convenient formulation for two reasons. First, after any history h^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 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. Second, the constraint that the contract prevent lapsation can be viewed as saying that the consumer's continuation utility starting in any period t when in health state λ_t cannot be less than in an optimal contract offered by an insurer who must break even, i.e., who has $S_t = 0$.

Thus, we begin by considering the period T contracting problem that arises when an insurer faces a consumer in health state λ_T and can sustain a loss of S_T:

$$\begin{array}{cc} \max_{c_T(\cdot)} & \int u(c_T(m_T))dF_T(m_T|\lambda_T) \\ s.t. & \int c_T(m_T)dF_T(m_T|\lambda_T) \leq S_T + y_T - \mathbb{E}[m_T|\lambda_T];\ \gamma_T \end{array} \tag{4}$$

In a long-term contract that breaks even overall, a positive subsidy $S_T > 0$ is enabled by previous front-loading of premiums – i.e., premiums that exceeded the insurer's expected medical expenses.[10] For the firm that initially contracted with the consumer in period 1, S_T may be positive, while for firms seeking to induce the consumer to lapse and sign with them, $S_T = 0$. We denote the value of this problem by $V_T(\lambda_T, S_T)$.

For periods $t < T$, we then consider the problem – whose value we denote by $V_t(\lambda_t, S_t)$ – that arises if a firm faces a consumer in health state λ_t and can sustain, going forward, a (discounted expected) loss of S_t:

$$\begin{array}{cll} \max_{c_t(\cdot), S_{t+1}(\cdot)} & & \int [u(c_t(m_t, \lambda_{t+1})) + \delta_t V_{t+1}(\lambda_{t+1}, S_{t+1}(m_t, \lambda_{t+1}))] dF_t(m_t, \lambda_{t+1}|\lambda_t) \\ s.t. & \text{(i)} & \int [c_t(m_t, \lambda_{t+1}) + \delta S_{t+1}(m_t, \lambda_{t+1})] dF_t(m_t, \lambda_{t+1}|\lambda_t) \leq S_t + y_t - \mathbb{E}[m_t|\lambda_t] \\ & \text{(ii)} & V_{t+1}(\lambda_{t+1}, S_{t+1}(m_t, \lambda_{t+1})) \geq V_{t+1}(\lambda_{t+1}, 0) \text{ for all } (m_t, \lambda_{t+1}) \end{array} \tag{5}$$

In this problem, the firm allocates resources to current consumption $c_t(\cdot)$ and to supporting future insurance through subsidies $S_{t+1}(\cdot)$, both of which are functions of the period t medical expense realization m_t and continuation state λ_{t+1} [as well as, implicitly, the state (λ_t, S_t)]. The first constraint is the break-even constraint for the firm going forward, including the subsidy S_t at the start of period t. The second constraint is the "lapsation constraint," which says that the consumer's utility going forward can never fall below the utility $V_{t+1}(\lambda_{t+1}, 0)$

[10] In particular, if there is front loading prior to period T, then a contract that breaks even overall will have $\mathbb{E}[S(\lambda_T)|\lambda_1] > 0$.

that a rival insurer (who has no subsidy – i.e., who must break even) can provide. Since the value that can be provided to the consumer is increasing in the subsidy, this constraint can equivalently be written as $S_{t+1}(m_t, \lambda_{t+1}) \geq 0$ – i.e., the optimal contract can never involve strictly positive discounted continuation profits (a negative "subsidy") for the insurer, since if it did a rival insurer could offer the consumer a greater continuation utility while still earning a strictly positive expected discounted profit.

Our main result, which can be viewed as a generalization of Harris and Holmstrom (1982), and which we establish in the Appendix, is:[11]

Proposition 1 *The equilibrium contract in a competitive market with one-sided commitment for a consumer with income path $y = (y_1, ..., y_T)$ and who cannot borrow is characterized by the consumption guarantees offered in the first period of a contract starting in period t with health state λ_t, $c_t^y(\lambda_t)$. The consumer who agrees to a contract in period 1 is fully insured against within-period medical expense risk, and enjoys in each period t following health state history $(\lambda_1, ..., \lambda_t)$ the certain consumption $\max_{\tau \leq t} c_\tau^y(\lambda_\tau)$. The levels $\{c_t^y(\lambda_t)\}$ lead the insurers to break even in expectation and consumers have no incentive to save after agreeing to the contract.*

Proof. In Appendix A. ■

Proposition 1 says that the equilibrium contract with one-sided commitment can be viewed as offering the consumer in period 1 an initial minimum guaranteed consumption level $c_1^y(\lambda_1)$. This minimal guarantee is bumped up in later periods $t > 1$ to a new, higher guarantee $c_t^y(\lambda_t) > c_1^y(\lambda_1)$ to match the outside market whenever the consumer's health state λ_t is sufficiently good that an outside firm could offer a better contract (i.e., a higher consumption guarantee) to the consumer and break even. The equilibrium contract provides full within-period insurance for the consumer (i.e., consumption that is independent of m_t), and partial insurance against reclassification risk, as consumers who have experienced sufficiently bad health states leading up to a given period t [such that $c_\tau^y(\lambda_\tau) \leq c_1^y(\lambda_1)$ for all $\tau < t$] all enjoy the same level of consumption regardless of differences in their health states. The extent of this partial insurance is a function of the consumer's initial health state λ_1, income path y, and risk preferences, 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.

[11] 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, show that the optimal contracts are self-selective (Section 2.3), and introduce switching costs (Sections 2.4 and 6.2).

The optimal contract in Proposition 1 specifies the consumer's consumption levels and prevents lapsation. Note, however, that the same outcome can alternatively be achieved by means of a *guaranteed renewable contract* from which the consumer may lapse. Specifically, the consumer is given the option to renew at the guaranteed renewable premium path $p^y \equiv (p_1^y, ..., p_T^y)$ where $p_t^y = y_t - c_1^y(\lambda_1)$ for $t = 2, ...T$, provided that he/she has always renewed in the past.[12] With this contract, some consumers, who have a sufficiently good health state λ_t, may choose not to renew in a given period t, instead signing a contract with a new insurer that offers premium path $p_\tau^y = y_\tau - c_t^y(\lambda_t)$ for $\tau \geq t$, where $c_t^y(\lambda_t) > c_1^y(\lambda_1)$. 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.[13,14]

2.3 Unobserved Income Profiles

The analysis above assumed that a consumer's lifetime income path $y = (y_1, ..., y_T)$ is known by both the consumer and all insurers. In practice, this is unlikely to be the case. In this subsection we show that insurers' failure to possess this information does not pose an obstacle to the use of optimal dynamic contracts. Specifically, we show that if offered the optimal contracts derived above, presented as the guaranteed renewable premium contracts described above, consumers will self-select, choosing the optimal contract for their income path.[15]

Specifically, define a "y-type" to be an individual with lifetime income path $y = (y_1, ..., y_T)$. We assume that there is a set Y of y-types in the market. As above, a guaranteed renewable

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

[13] We discuss in Section 8 the difference between this guaranteed renewable policy and that of Pauly, Kunreuther, and Hirth (1995).

[14] 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 λ_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 $S_2^*(m_1, \lambda_2)$ from the solution to problem (5) for $t = 1$. The premium in period 1 would equal $y_1 - c_1^*(\lambda_1)$, while the period 2 premium would equal $y_2 - \max_{\tau \leq 2} c_\tau^*(\lambda_\tau) + S_2^*(m_1, \lambda_2)$. This is exactly the amount that is needed to buy a long-term policy in period 2 that offers the consumer a consumption guarantee equal to $\max_{\tau \leq 2} c_\tau^y(\lambda_\tau)$. 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 in to 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.

[15] 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.

premium path is a $p = (p_1, ..., p_T)$ that allows the consumer to continue coverage in period t paying premium p_t provided that she has not previously lapsed. The optimal guaranteed renewable contract for a given known y-type is, as above, denoted by the path p^y, a path that keeps consumption constant as income changes from year to year according to path $y = (y_1, ..., y_T)$.

Our result is:

Proposition 2 *If an insurer offers the menu of guaranteed premium paths $\{p^y\}_{y \in Y}$, where $p^y \equiv \{y_t - c_1^y(\lambda_1)\}_{t=1}^T$ for each y-type in Y, each y-type optimally chooses contract p^y. That is, the menu $\{p^y\}_{y \in Y}$ is self-selective.*

Proof. In Appendix B. ■

2.4 Switching Costs

Recent evidence suggests that consumers may exhibit substantial inertia in their health insurance choices [see, e.g., Handel (2013)]. In Section 6.2 we 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. We show in Appendix A that the key change this introduces is that the lapsation constraint in the period t problem (5) becomes $S_{t+1}(m_t, \lambda_{t+1}) \geq -\sigma$; that is, the insurer can now earn positive discounted expected continuation profits up to σ without causing the consumer to lapse. Nonetheless, the basic structure of an optimal contract is unchanged when switching 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.

3 Data and Parameter Estimates

We investigate positive and normative outcomes for each type of contracting relationship. To predict equilibrium contracts and welfare under each regime we need four basic ingredients: i. the distribution of health risks for each population cohort, ii. the transitions across health risks (states) as individuals age, iii. preferences towards risk, and iv. income paths.

We use detailed administrative data on the health insurance choices and medical utilization of employees (and their dependents) at a large U.S.-based firm over the period 2004 to 2009. These proprietary panel data include information on employee plan choices, and detailed, claim-level employee (and dependent) medical expenditure and utilization information. Overall, the data include 11,253 employees and 9,710 dependents, implying a total of

20,963 covered lives. For more information and descriptives statistics see Handel, Hendel and Whinston (2015).

The sample used in our analysis includes individuals between the ages of 25 and 65 who are present over a given two-year period in our data (two years are needed to estimate health state transitions for those individuals). The sample displays ages close to uniformly spread between 25 and 65, is 45% male, and has incomes dispersed over the full range we can measure from Tier 1 (less than $41,000) to Tier 5 (greater than $176,000).

The most essential part of the data is the available information on diagnostics (ICD-9 codes) of each individual in the sample. The diagnostic codes as well as other demographics are fed 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.[16] The output is an index that represents the health state of each individual in the population. Since the ACG is used by insurers in their underwriting processes, our empirics are based on the same information about risks that market participants (insurers) have.[17]

3.1 Health Risk Distribution

We denote the ACG index by λ and we refer to λ_{it} as individual i's "health state" at time t. As the ACG score measures expected medical expenses, Assumption 1 in the previous section holds by construction, though we verify that the higher predicted expenditures do imply higher actual expenditures and vice-versa.

The distribution of λ_{it} for each age t represents the distribution of health state in each population cohort. Table 1 presents the decomposition of health expenses between the predictable component λ, and the deviation around this expectation. The former reflects heterogeneity and the potential source of reclassification risk, while the latter captures the gains from spot insurance (conditional on λ).[18] For each age cohort, the first two columns show the overall mean expense and the overall standard deviation around this mean. The last two columns then decompose this standard deviation into the standard deviation of $\mathbb{E}[m|\lambda]$ and the average standard deviation of medical expenses around $\mathbb{E}[m|\lambda]$. The table illustrates

[16] 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.

[17] 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 health outcomes. Instead our measure of risk is based on professional diagnostics (the ACG index). The distribution of risk is data, as opposed to being inferred from choice.

[18] One of the market configurations we consider, the ACA, does not entail full event insurance. Computing welfare without full event insurance requires as an input the distribution of health outcomes (expenses) conditional on risk type, rather than just its mean. See Handel, Hendel, and Whinston (2015).

Sample Total Health Expenditure Statistics				
Ages	Mean	S. D.	S. D. of $E[m\|\lambda]$	S. D. around $E[m\|\lambda]$
All	6,099	13,859	6,798	9,228
25-30	3,112	9,069	4,918	5,017
30-35	3,766	10,186	5,473	5,806
35-40	4,219	10,753	5,304	6,751
40-45	5,076	12,008	5,942	7,789
45-50	6,370	14,095	6,874	9,670
50-55	7,394	15,315	7,116	11,092
55-60	9,175	17,165	7,414	13,393
60-65	10,236	18,057	7,619	14,366

Table 1: Sample statistics for total health expenditures for (i) the entire sample used in our equilibrium analysis and (ii) 5-year age buckets within that sample. "Mean" column reports the average medical cost for the relevant group; "S.D." reports its standard deviation; "S.D. of $E[m|\lambda]$" and "S.D. around $E[m|\lambda]$" columns decompose the overall standard deviation into the parts related to the observable health state and the residual variation.

that both within-period insurance for expenditure risk and longer-run insurance for health state transitions are important aspects of consumer risk protection in this market. The large standard deviation in $\mathbb{E}[m|\lambda]$ conditional on age suggests that there could be significant reclassification risk in spot contracting environments, as well as well as meaningful adverse selection if community rating is introduced (as under the Affordable Care Act), something that is shown in Handel, Hendel, and Whinston (2015).[19]

3.2 Health State Transitions

The second key input into our empirical analysis are the health state transitions over time. Once we have λ_{it} for every individual in the sample, we estimate the probabilistic health state transition process as follows. First, we divide the λ into seven mutually exclusive and exhaustive bins that each contain one-seventh of the final sample. Table 2 presents the proportion of individuals in each age group in each of these seven health state categories. We then compute a separate transition matrix for each five-year age group (e.g., transitions within cohort 25-30) using the actual transitions of consumers within each age range. We

[19] In Handel, Hendel, and Whinston (2015) we analyze risk-adjustment transfers designed to mitigate adverse selection in markets with community rating, such as the state-based ACA exchanges. As shown in Brown et al. (2014), in practice these policies have helped mitigate adverse selection, but not eliminated it as a problem.

Health State by Age							
Age	1 (Healthy)	2	3	4	5	6	7 (Sick)
25-30	0.49	0.19	0.14	0.07	0.04	0.03	0.04
30-35	0.41	0.18	0.13	0.08	0.06	0.06	0.07
35-40	0.27	0.30	0.13	0.06	0.09	0.07	0.09
40-45	0.19	0.28	0.16	0.09	0.12	0.08	0.10
45-50	0.01	0.15	0.32	0.15	0.13	0.12	0.12
50–55	0.00	0.10	0.25	0.19	0.15	0.16	0.15
55-60	0.00	0.01	0.01	0.25	0.24	0.28	0.22
60-65	0.00	0.00	0.00	0.18	0.24	0.26	0.31

Table 2: Health state by age, where consumers are divided into 7 bins of their predicted medical spending (determined by their Johns Hopkins ACG predictive score) for the year ahead.

limit the partition of the health space to seven bins, and the age groups to five-year bins, to guarantee meaningful support to estimate each probability in these transition matrices. We use the eight 7-by-7 transition matrices for the five-year age bins from 25-65 as the foundation for modeling health state persistence and transitions over time.

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.[20]

Tables 3 and 4 present these transition matrices for ages 30-35 and 50-55 respectively. Entries along the diagonal of each matrix reflect health state persistence, while off-diagonal elements reflect health state changes. For example, 72% of consumers aged 30-35 who are in the healthiest state bin in one year ($\lambda_{it} = 1$) are estimated to stay in that category for the next year. Only 11% of these consumers begin the next year in one of the four sickest bins ($\lambda_{it} \in \{4, 5, 6, 7\}$). On the other hand, 57% of consumers who begin the first year in the sickest health state bin ($\lambda_{it} = 7$) begin the next year in one of the two sickest bins ($\lambda_{it} \in \{6, 7\}$). Though the distributions of health are different for 50-55 year olds, health states show similar persistence.

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 starting at age 30 for consumers in different health states at age 30. Table 6 shows the same for consumers starting at age 45. The tables show that while there is significant persistence, much of it is dissipated in 10 years, and to a large extent after 5 years.[21]

[20] 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.

Health State by Age							
				λ_{t+1}			
λ_t	1	2	3	4	5	6	7
$\lambda_t = 1$	0.72	0.13	0.05	0.05	0.02	0.01	0.03
$\lambda_t = 2$	0.35	0.25	0.12	0.11	0.04	0.03	0.11
$\lambda_t = 3$	0.15	0.23	0.19	0.15	0.10	0.08	0.10
$\lambda_t = 4$	0.20	0.08	0.12	0.24	0.18	0.12	0.08
$\lambda_t = 5$	0.10	0.10	0.05	0.20	0.20	0.20	0.15
$\lambda_t = 6$	0.16	0.11	0.14	0.11	0.08	0.22	0.19
$\lambda_t = 7$	0.11	0.11	0.07	0.04	0.11	0.20	0.37

Table 3: Health state transitions from one year to the next, for 30-35 year olds.

Health State by Age							
				λ_{t+1}			
λ_t	1	2	3	4	5	6	7
$\lambda_t = 1$	0.67	0.15	0.10	0.02	0.02	0.01	0.03
$\lambda_t = 2$	0.25	0.37	0.20	0.09	0.04	0.02	0.04
$\lambda_t = 3$	0.09	0.21	0.21	0.20	0.12	0.10	0.08
$\lambda_t = 4$	0.10	0.19	0.26	0.12	0.10	0.19	0.05
$\lambda_t = 5$	0.09	0.19	0.14	0.15	0.10	0.19	0.15
$\lambda_t = 6$	0.00	0.09	0.13	0.09	0.19	0.23	0.28
$\lambda_t = 7$	0.03	0.10	0.10	0.10	0.21	0.16	0.29

Table 4: Health state transitions from one year to the next, for 50-55 year olds.

Age 30 health state	Present value of expected expenses at various ages			
	30	31-35	36-40	31-65
1	1,131	16,888	20,066	135,132
2	2,290	22,184	20,482	140,884
3	3,780	25,155	20,761	144,159
4	3,975	25,752	20,810	144,809
5	5,850	29,228	20,960	148,450
6	10,655	28,854	20,853	147,950
7	18,554	33,366	20,988	152,618

Table 5: The table shows present value of expected health expenses for different periods (age 30, ages 31 to 35, ages 36-40 and 31 to 65), for an individual who at age 30 was in the respective health state.

Age 40 health state	Present value of expected expenses at various ages			
	40	41-45	46-50	41-65
1	1,131	19,587	24,317	119,397
2	2,290	23,827	24,753	124,126
3	3,780	28,023	25,096	128,705
4	3,975	28,993	25,127	129,710
5	5,850	30,851	25,236	131,690
6	10,655	33,376	25,316	134,304
7	18,554	38,403	25,552	139,532

Table 6: The table shows present value of expected health expenses for different periods (age 40, ages 41 to 45, ages 46-50 and 41 to 65), for an individual who at age 40 was in the respective health state.

3.3 Risk Preferences

The final ingredient is consumer's risk aversion, i.e., the degree to which consumption smoothing over different states of the world is valued by consumers. To this end, we use the risk preferences estimated from our data 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:

[21] In Appendix C we examine how the level of health state persistence affects welfare under optimal dynamic contracts relative to our benchmarks. Two points to note are: (i) reclassification risk is still present even when there is no persistence in the transitions of λ, since λ_{t+1} is observed in period t (in a sense, there is always at least one-year of persistence of information regarding health risks), and (ii) with perfect persistence of λ there is no benefit from optimal dynamic contracts with one-sided commitment as there would be no reclassification risk to insure once a consumer's age 25 health state is realized.

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

where $c = y - p - m$ is consumption, which equals income net of health expenses (premium payments p plus out-of-pocket medical expenses m) and γ is the risk aversion parameter.[22]

Handel, Hendel, and Whinston (2015) estimate the distribution of γ which fits the observed choices, given the choice sets available each year of the sample. The mean estimated risk-aversion level is 0.000428 and the standard deviation in this mean from observed heterogeneity is 0.00001. The estimates fall within the range reported in the literature. See Handel, Hendel, and Whinston (2015) for further discussion of these risk preference estimates.

In the next sections we use the median risk aversion estimate $\gamma = 0.0004$. We also consider the robustness of our conclusions with respect to the degree of risk aversion in Section 6.1.

3.4 Income Paths

The shape of the optimal contract depends on a consumer's income path. Firms offer different contracts to consumers with different income paths to maximize their lifetime expected utilities conditional on breaking-even and the lapsation constraint. In our analysis, we consider the income paths of managers at the firm which are steepest, those of non-managers which are flatter, and a baseline of a completely flat net income path.[23] The latter neutralizes consumption smoothing motives, 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. The manager income path is the steepest and highest dashed curve depicted in Figure 1, while non-managers' income path is represented by the flatter curve comprised of dashes and dots. The bold dashed curve is a proportionally scaled-down managers' income path 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 path. As can be seen in the figure, both managers and non-managers have income paths that grow more quickly than expected medical expenses, leading to steeper paths than the flat net income curve.

[22] In the model, we assume that consumers have zero price elasticity of demand for health care services, which is why full insurance is efficient in our setup. See Handel, Hendel, and Whinston (2015) for an extended discussion of that assumption.

[23] The manager and non-manager income paths we use here are estimated in Handel, Hendel, and Whinston (2015).

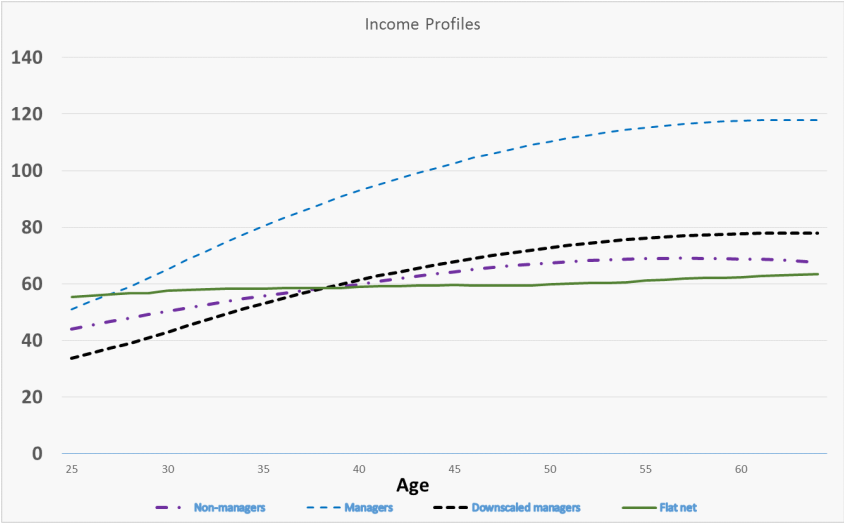


Figure 1: Income paths for managers, non-managers, proportionally "downscaled" managers, and flat net income.

4 Computation of Optimal Dynamic Contracts

Given estimated values of $f(\cdot)$ and $u(\cdot)$, income profile y, and a value for δ, we derive the contracts described in Proposition 1 recursively. We first derive, for each possible λ_T, the last period consumption levels $c_T^y(\lambda_T) = y_T - \mathbb{E}[m_T|\lambda_T]$ that would be offered to a consumer in state λ_T by competitive firms. We then look at each possible state λ_{T-1} in period $T-1$. We find $c_{T-1}^y(\lambda_{T-1})$ by doing a binary search over possible values for the consumption guarantee 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 in which $c_T^y(\lambda_T) > c_{T-1}$. We then continue backward in this fashion, with the transition matrix 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, for each period t, we denote by C_t^y the $(T-t+1)\times 7$ matrix of first-period consumption guarantees whose (τ, s) element for $\tau > t$ and $s \in \{1, ..., 7\}$ is $c_\tau^y(\lambda_\tau = s)$, a consumption guarantee that breaks even given the future guarantees described in Proposition 1 (and, themselves, contained in C_{t+1}^y). We start at $t = T$ where, as noted above, $c_T^y(\lambda_T = s) = y_T - \mathbb{E}[m_T|\lambda_T = s]$. This gives us C_T^y. We then proceed iteratively backwards, deriving C_t^y given C_{t+1}^y and the transition probabilities. Specifically, C_t^y adds an additional row to C_{t+1}^y; each element (t, s) of this row is the consumption guarantee $c_t^y(\lambda_t = s)$. We derive

$c_t^y(\lambda_t = s)$ by doing a grid search to find the (unique) value $\overline{c}$ such that

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

where $P_\tau(s|\lambda_t, C_{t+1}^y, \overline{c})$ is the probability that, starting in health state $\lambda_t = s$ in period t, the health state transitions $(\lambda_{T+1}, ..., \lambda_\tau)$ from period t to period τ are such that $c_{t'}^y(\lambda_{t'}) \leq \overline{c}$ for all $t' \in \{t+1, ..., \tau\}$.[24] 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 (with the summation signs) is the expected continuation profit and uses the fact that starting at the first instance at which the consumer's health state λ_τ is such that $c_\tau^y(\lambda_\tau) > \overline{c}$ (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^y, whose first row gives the initial consumption guarantees offered in period 1 to consumers in each of the seven possible period 1 health states.

5 Results

5.1 Front-loading and Reclassification Risk

We describe first the extent of front-loading in the equilibrium contracts, and then consider the amount of remaining reclassification risk. We begin by considering the contract for a population with flat net income (corresponding to the solid curve in Figure 1), an income path that creates no borrowing or savings motive.

In our context, a contract specifies a premium and resulting consumption level for each possible history of states at each age from 25 to 65. There are too many histories and concurrent premiums 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, actuarial costs, and front-loading are presented in Table 7. In all but the worst state, premiums are larger than actuarial costs: consumers front-load premiums to transfer utility to future states with negative health shocks (consumers in the worst state have no desire to transfer utility into the future given their high current needs). For the healthiest consumer at the beginning of the year, the premium is \$2,750 despite average costs of only \$1,131.

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

[24] These probabilities are computed using C_{t+1}^y and the transition matrices.

First-Year Equilibrium Contract Terms: Flat Net Income							
	λ_1						
	1	2	3	4	5	6	7
Premium	2,750	4,155	6,008	6,130	8,885	11,890	18,554
First-Year Costs	1,131	2,291	3,780	3,975	5,850	10,655	18,554
Front-Loading	1,619	1,864	2,228	2,155	3,035	1,235	0
Consumption	52,550	51,145	49,292	49,170	46,415	43,410	36,746

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

front-loading depends on both the current state and also on the implications of the current state for future health. While the healthier type can afford the most front-loading, she 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. It is important to note that front-loading in this example is not driven by saving motives, since net income is flat. Front-loading is done to enable inter-temporal insurance against poor risk realizations.

Tables 8 and 9 present second-period premiums and consumption, respectively, reflecting the 7 x 7 possible histories up through the beginning of that period (health state at the beginning of period 1 crossed with health state at the beginning of period 2). 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 in period 1, all other second-year states have the same consumption of \$52,550, an amount equal to her first-year consumption. The lapsation constraint does not bind for this consumer when $\lambda_2 > 1$ because the first period front-loaded amount suffices to make outside offers less attractive than the current consumption guarantee.

The lapsation constraint binds for more and more second-year states the sicker consumer was at the start of the contract. For consumers initially in the sickest health bin ($\lambda_1 = 7$), all second-year health states involve different consumption levels that also differ from the period 1 consumption level: long-run contracts cannot provide any insurance against reclassification risk in year 2 for this consumer as her first-year needs were so great as to preclude any front loading. For this consumer, the long-run contract continuation in year 2 simply matches the best contract she could get on the market given her second-year health state.

It is difficult to display premiums beyond the first couple of years, as there are 7^{t-1}histories to consider in year t. For example, by year 10 there are over 40 million histories. Table 10 shows statistics associated with premiums for the first ten contract years including the mean and maximum front-loading across possible histories. Specifically, the table shows for each

Second-Year Equilibrium Premiums: Flat Net Income								First-Year
				λ_2				Premium
λ_1	1	2	3	4	5	6	7	
1	2,943	3,300	3,300	3,300	3,300	3,300	3,300	2,750
2	2,943	4,302	4,705	4,705	4,705	4,705	4,705	4,155
3	2,943	4,302	6,090	6,206	6,558	6,558	6,558	6,008
4	2,943	4,302	6,090	6,206	6,680	6,680	6,680	6,130
5	2,943	4,302	6,090	6,206	8,955	9,434	9,434	8,885
6	2,943	4,302	6,090	6,206	8,955	11,919	12,440	11,890
7	2,943	4,302	6,090	6,206	8,955	11,919	18,554	18,554

Table 8: First- and second-year premiums in the equilibrium long-run contract for a flat net income path, as a function of the period 1 and period 2 health states.

Second-Year Equilibrium Consumption: Flat Net Income								First-Year
				λ_2				Consumption
λ_1	1	2	3	4	5	6	7	
1	52,905	52,550	52,550	52,550	52,550	52,550	52,550	52,550
2	52,905	51,545	51,145	51,145	51,145	51,145	51,145	41,145
3	52,905	51,545	49,758	49,642	49,292	49,292	49,292	49,292
4	52,905	51,545	49,758	49,642	49,170	49,170	49,170	49,170
5	52,905	51,545	49,758	49,642	46,893	46,415	46,415	46,415
6	52,905	51,545	49,758	49,642	46,893	43,929	43,410	43,410
7	52,905	51,545	49,758	49,642	46,893	43,929	37,294	36,746

Table 9: First- and second-year consumption in the equilibrium long-run contract for a flat net income path, as a function of the period 1 and period 2 health states.

Second-Year Equilibrium Premiums										
	Contract Year									
	1	2	3	4	5	6	7	8	9	10
Mean Front-Loading	1,702	1,682	1,758	1,834	1,919	2,280	2,461	2,536	2,568	2,469
Max. Front-Loading	3,035	3,105	3,174	3,324	3,236	3,324	3,623	3,781	3,859	3,438

Table 10: Mean and maximum front-loading for each of the first ten years of the equilibrium long-run contract.

First-Year Equilibrium Contract Term: Downscaled Managers							
	λ_1						
	1	2	3	4	5	6	7
Premium	1,678	3,165	5,256	5,449	8,245	11,593	18,554
First-Year Costs	1,131	2,291	3,780	3,975	5,850	10,655	18,554
Front-Loading	547	874	1,476	1,474	2,395	938	0
Consumption	32,022	30,535	28,244	28,051	25,255	21,907	14,946

Table 11: First-year contract terms in the equilibrium long-run contract for a downscaled manager income path, showing first-year premium, expected costs, the extent of contract front-loading, and consumption levels.

of the first ten years the mean front-loading (averaging across all possible histories, weighted by the probability the histories). The second row shows the maximum possible front-loading in each year, across these ten periods. Mean front-loading increases over time from \$1,702 in year 1 to \$2,469 in year 10. Maximum possible front-loading increases slightly over time.

5.2 Effects of Income Paths

The equilibrium contracts offered depend crucially on a consumer's rate of income growth over her lifetime. When income is relatively low early in life, front-loading is quite costly and undesirable in the absence of well-functioning capital markets. In our analysis, "managers," who are higher-level employees, have steeper income paths as they age, relative to non-managers, whose net income paths are flatter.

Table 11 presents first-period contract characteristics for "downscaled managers," and is the analog to Table 7 for flat net income. Recall that, as shown in Figure 1, a downscaled manager income path proportionally scales down the income of a manager to match the net present value of a non-manager's lifetime income path. 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. Essentially, the rapidly rising income makes paying extra for insurance early quite costly.

Second-Year Equilibrium Premiums: Downscaled Managers								First-Year
				λ_2				Premium
λ_1	1	2	3	4	5	6	7	
1	1,699	3,204	3,524	3,524	3,524	3,524	3,524	1,678
2	1,699	3,204	5,011	5,011	5,011	5,011	5,011	3,165
3	1,699	3,204	5,289	5,478	7,102	7,102	7,102	5,256
4	1,699	3,204	5,289	5,478	7,295	7,295	7,295	5,449
5	1,699	3,204	5,289	5,478	8,279	10,090	10,090	8,245
6	1,699	3,204	5,289	5,478	8,279	11,615	13,439	11,593
7	1,699	3,204	5,289	5,478	8,279	11,615	18,554	18,554

Table 12: First- and second-year premiums in the equilibrium long-run contract for a downscaled manager income path, as a function of the period 1 and period 2 health states.

Second-Year Equilibrium Consumption: Downscaled Managers								First-Year
				λ_2				Consumption
λ_1	1	2	3	4	5	6	7	
1	33,833	32,328	32,008	32,008	32,008	32,008	32,008	32,008
2	33,833	32,328	30,521	30,521	30,521	30,521	30,521	30,521
3	33,833	32,328	30,243	30,054	28,430	28,430	28,430	28,430
4	33,833	32,328	30,243	30,054	28,237	28,237	28,237	28,237
5	33,833	32,328	30,243	30,054	27,253	25,441	25,441	25,441
6	33,833	32,328	30,243	30,054	27,253	23,917	22,093	22,093
7	33,833	32,328	30,243	30,054	27,253	23,917	16,978	15,132

Table 13: First- and second-year consumptions in the equilibrium long-run contract for a downscaled manager income path, as a function of the period 1 and period 2 health states.

Tables 12 and 13 show second-year premiums and consumption levels for downscaled managers as a function of different health histories. Though front-loading is much more limited, the year 2 states in which the lapsation constraint binds, conditional on the initial year 1 health state, are quite similar to those of consumers with flat net income profiles.

5.3 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 path 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, ex ante.[25] Specifically, we compare the certainty

[25] The consumer arrives at age 25 with the health distribution given in the first row of Table 2.

equivalent of optimal dynamic contracts with one-sided commitment, denoted by CE_D, to (see Section 2.1 for formal definitions):

(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 Saving/No Borrowing constrained first best, in which risk is fully insured in each period but no borrowing is possible, denoted C^*_{NBNS};

(iv) The certainty equivalent of an ACA-like market exchange, denoted CE_{ACA}.

We also compare optimal dynamic contracts with one-sided commitment to an alternative market design used to cope with reclassification risk, ACA-like exchanges. These markets have year-to-year contracts, but are community rated: they thus fully insure premium reclassification risk, at the potential cost of adverse selection for within-year coverage. In Handel, Hendel and Whinston (2015), adverse selection in this market is found to be substantial, as it fully unravels with all consumers in minimum coverage (60%) plans in equilibrium.

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 coverage purchase, and (v) insurers that offer coverage of specific actuarial values, with a minimum plan that covers 60% of spending for an average population. As shown in Handel, Hendel, and Whinston (2015), this market eliminates reclassification risk, but induces adverse selection that reduces within-period event insurance. The market unravels to minimum coverage (60%). Insurees end up responsible for 40% of their costs on average. Because of the partial coverage consumers face some future risk, not from increased premiums, but rather from out-of-pocket expenses. We denote the certainty equivalent of this outcome by CE_{ACA}. Notice that C^*_{NBNS} is the same as this 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.

To ease comparisons across different income paths, all profiles we consider have the same net present value of income, except for the manager's profile when it is not downscaled. In the first best where consumers are fully insured and income is fully smoothed over time,

all income paths (except managers without downscaling) receive a certainty equivalent of $53,666 (reflecting the present value of income minus expected medical expenses).

Column (1) of Table 14 confirms C^* is $53,666 for all income profiles (except for managers). Column (2) shows welfare outcomes under spot contracts, with no protection for reclassification risk. As column (6) shows, the gap between the two is highest for relatively steep income profiles (69% for downscaled managers), as the negative shocks early in life in spot markets are relatively more costly for individuals with rapid income growth. However, the welfare loss from spot contracts is substantial even for flatter income profiles (16% for flat net income).

Column (3) presents the certainty equivalent for dynamic contracts with one-sided commitment, CE_D. As expected CE_D lies between C^* and CE_{SPOT}. Column (7) shows the fraction of the welfare gap between the first best and spot contracts that these dynamic contracts close, $\frac{CE_D - CE_{SPOT}}{C^* - CE_{SPOT}}$. For flatter income profiles, dynamic contracts almost fully insure reclassification risk and are quite close to first best: in the case of flat net income 87.7% of this gap is recovered by dynamic contracts. In contrast, since consumers with steeper income profiles dislike front-loading, dynamic contracts only recover a small portion of this welfare loss for steeper income profiles (10.8% for downscaled managers).

Column (4) shows C^*_{NBNS}, the constant consumption equivalent of the constrained first-best outcome that does not allow for intertemporal consumption smoothing. For situations with rising net income, a more telling measure of the performance of dynamic contracts is $\frac{CE_D - CE_{SPOT}}{C^*_{NBNS} - CE_{SPOT}}$, which is shown in column (8).[26] This statistic removes the benefits of consumption smoothing from saving and borrowing that contracts with two-sided commitment deliver. This is potentially quite important, since saving and borrowing on their own can greatly improve utility for steep net income paths. This statistic reflects the proportion of the gap between spot contracting and the first best closed by optimal dynamic contracts that is unrelated to smoothing consumption in the first best. The progress made by dynamic contracts relative to this constrained first best is a good deal higher than when using C^* as a benchmark. Dynamic contracts get downscaled managers 38.1% of the way to the constrained first best, instead of 10.8%, and those with flat net income get 89.5% of the way to the constrained first best. Nonetheless, the analysis makes clear that dynamic contracts are quite effective when net income paths are fairly flat, but significantly less so with rising income paths.

Comparing columns (3) and (5) in Table 14 reveals that the ACA environment is preferred to the dynamic contracting environment for non-managers, managers, and downscaled

[26] The fact that, for flat income, dynamic contracts (C_D) do better than C^*_{NBNS} [see column (8)] is because with flat income, which generates a declining net income path, the first best involves moving resources later through savings, which is achievable with dynamic contracts, but not in the no-borrowing/no-savings regime.

	Certainty Equivalent ($1,000s)					Gains from Long-term contracts		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Income	C^*	CE_{SPOT}	CE_D	C^*_{NBNS}	CE_{ACA}	$\frac{C^*-CE_{SPOT}}{CE_{SPOT}}$	$\frac{CE_D-CE_{SPOT}}{C^*-CE_{SPOT}}$	$\frac{CE_D-CE_{SPOT}}{C^*_{NBNS}-CE_{SPOT}}$
Flat ($58.9K)	53.67	45.88	53.46	52.72	49.69	17.0%	97.3%	110.8%
Flat net	53.67	46.27	52.77	53.53	51.30	16.0%	87.7%	89.5%
Non-mngr	53.67	40.73	44.10	47.39	46.25	31.8%	26.0%	50.6%
Manager	84.00	50.32	51.77	56.08	55.09	66.9%	4.3%	25.2%
Downs Mngr	53.67	31.74	34.10	37.93	36.84	69.1%	10.8%	38.1%

Table 14: Long-run welfare results showing the certainty equivalent consumption of different insurance institutions under a discount factor of 0.975 for consumers with median estimated constant absolute risk aversion equal to 0.0004.

managers, whose incomes rise over time. For these individuals, the desire to front-load when young and income is low is limited, which reduces the benefits from dynamic contracts. Interestingly, individuals with flat net income profiles prefer dynamic contracts to the ACA environment, so the optimal market design depends on which of several types of income paths individuals have.

Finally, note that the gap between CE_{ACA} and C^*_{NBNS} reflects the cost of adverse selection in the exchanges relative to the constrained first-best. Both regimes involve community rating, the first with consumers buying plans with 60% actuarial value (due to adverse selection) and the second with consumers in full insurance. This cost of adverse selection ranges from roughly $1,000 per year (downscaled managers) to roughly $2,200 per year for flat net income.

6 Extensions and Robustness

6.1 Risk Aversion

So far our analysis has used the median 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 health insurance choice, might not reflect the larger stakes associated with reclassification risk.[27] The scale of losses from reclassification risk is larger in our environment than in the setting in which risk preferences were estimated.

[27] See Rabin (2000) for a discussion of issues with CARA and scaling of gambles.

	Certainty Equivalent ($1,000s)					Gains from Long-term contracts		
Income	C^*	CE_{SPOT}	CE_D	C^*_{NBNS}	CE_{ACA}	$\frac{C^*-CE_{SPOT}}{CE_{SPOT}}$	$\frac{CE_D-CE_{SPOT}}{E^*-CE_{SPOT}}$	$\frac{CE_D-CE_{SPOT}}{C^*_{NBNS}-CE_{SPOT}}$
Flat-net	53.67	52.47	53.62	53.67	52.85	2.3%	95.8%	95.8%
Non-mngr	53.67	50.91	51.70	51.84	51.35	5.4%	28.6%	85.0%
Manager	84.00	68.73	69.06	69.39	69.00	22.2%	2.2%	50.0%
Downs mngr	53.67	46.41	46.94	47.20	46.80	15.6%	7.3%	67.1%

Table 15: Long-run welfare results 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.

Table 15 presents the welfare comparisons for risk aversion of 0.00008, five times lower than that in our main analysis. To put the coefficients in perspective consider a lottery that assigns the loss associated with each of the 7 health states, with each of them having equal probability. (The expected health care costs are reported in Table 7.) Our 0.0004 risk aversion coefficient estimate implies a willingness to pay of $\$7,222$ to avoid the uncertainty associated with such risky prospect. Instead, the lower risk coefficient leads to a willingness to pay of $\$1,491$.

Lowering risk aversion substantially reduces the loss associated with reclassification risk (captured by the gap between C^*_{NBNS} and CE_{SPOT}). The loss is, on average across income paths, only 1.6%. For the higher 0.0004 risk aversion, the loss was above 13%.

While the loss from reclassification risk is lower, the gains from long-term contracting, as captured by how much of the gap is recaptured by optimal contracts with one-sided commitment, is larger. Optimal dynamic contracts come very close to eliminating the gap for all income profiles. It appears that most of the welfare loss is eliminated with tolerable levels of front-loading, while the remaining reclassification risk is not that harmful for the less risk averse consumer.

The lower risk aversion not only attenuates the cost of reclassification risk, increasing welfare under spot and dynamic contracts, but also lowers the losses from the partial coverage of medical costs under the ACA. However, for lower risk aversion, dynamic contracts are preferred to the ACA for all four income paths.

6.2 Switching Costs

Recent evidence from health insurance markets [Handel (2013)] points to substantial inertia in insurance choice. Switching costs have the potential to impact how dynamic contracts perform. Our basic model assumes that consumers lapse whenever they get a better offer. As

Switching Cost		Flat-net	Non-manager	Managers	Downs mngr
0		52.76	44.10	51.77	34.10
100		52.79	44.16	51.85	34.17
$1,000$		52.95	44.68	52.54	34.95
$5,000$		53.39	46.45	55.07	36.92
$10,000$		53.58	47.85	57.45	38.82
$50,000$		53.66	51.67	65.37	44.72
$100,000$		53.67	53.52	70.75	48.76
$500,000$		53.67	53.67	84.00	53.67
C^*		53.67	53.67	84.00	53.67
C^*_{NBNS}		53.53	47.39	56.08	37.93

Table 16: Long-run welfare (in \$1,000s) from optimal dynamic contracts (CE_D) under discount factor of 0.975 for different levels of switching costs, and income profiles, and a risk aversion of 0.0004.

discussed in Section 2.4 switching costs relax the lapsation constraints, which can enhance commitment and the welfare achievable with optimal dynamic contracts.

Table 16 shows the welfare achieved by dynamic contracts for different levels of switching costs and income profiles. As expected, this welfare is monotonic in the switching cost.

Relaxing the lapsation constraints enables the transfer of resources from good health states to less fortunate ones. Qualitatively, as switching costs increase, from 0 to infinity, welfare in the optimal dynamic contract with one-sided commitment approaches the first-best (two-sided commitment) level.

Notice that while it takes an unreasonably large switching cost to achieve a welfare close to first best, a switching cost of around \$5,000 delivers welfare close to CE_{ACA} for consumers with rising income paths, and a switching cost between \$5,000 and \$10,000 achieves welfare comparable to C^*_{NBNS}. The reason is that the commitment afforded by switching costs helps in two ways. By lowering lapsation, commitment allows for a reduction in reclassification risk. In addition, enhanced commitment enables the insurer to facilitate consumption smoothing across periods. However, consumption smoothing through borrowing requires a lot of commitment, especially when the income profile is steep. Thus, a substantial switching cost is necessary to achieve the first best. Instead achieving the welfare level C^*_{NBNS} requires more modest inertia, as no consumption smoothing is involved.

Table 17 shows first and second-year consumption levels for a flat net income profile and a switching costs of \$5,000 and \$1,000. It is interesting to note that for all second-year states without a binding lapsation constraint premiums are lower with a higher switching cost,

First and Second-Year Consumption: Switching Cost = 5,000								First Year
				λ_2				Consump
λ_1	1	2	3	4	5	6	7	
1	53,612	53,612	53,612	53,612	53,612	53,612	53,612	53,613
2	52,910	52,765	52,765	52,765	52,765	52,765	52,765	52,765
3	52,910	51,717	51,717	51,717	51,717	51,717	51,717	51,717
4	52,910	51,551	51,243	51,243	51,243	51,243	51,243	51,243
5	52,910	51,551	49,763	49,647	49,224	49,224	49,224	49,224
6	52,910	51,551	49,763	49,647	46,898	46,881	46,881	46,881
7	52,910	51,551	49,763	49,647	46,898	43,934	40,652	40,652
First and Second-Year Consumption: Switching Cost = 1,000								
1	53,088	53,088	53,088	53,088	53,088	53,088	53,088	53,088
2	53,010	51,733	51,733	51,733	51,733	51,733	51,733	51,733
3	53,010	51,651	50,001	50,001	50,001	50,001	50,001	50,001
4	53,010	51,651	49,901	49,901	49,901	49,901	49,901	49,901
5	53,010	51,651	49,863	49,747	47,162	47,162	47,162	47,162
6	53,010	51,651	49,863	49,747	46,998	44,301	44,301	44,301
7	53,010	51,651	49,863	49,747	46,998	44,034	37,758	37,759

Table 17: First and second-year consumptions for switching costs of $5,000 and $1,000 and flat net income.

while for second-year states with binding lapsation constraint premiums are lower. Namely, conditional on a history, higher switching costs enable transferring resources from the good to the bad states.

6.3 Transitions

One of the key ingredients in our simulation are the health state transitions. Tables 3 and 4, while reasonable, display some violations of Assumption A2. The violations probably reflect the sparsity of individuals in each cell. To study the robustness of our results to the transition matrices, we present equilibrium contracts using the transitions from a larger employer with over 50,000 insured lives.[28]

Table 18 shows welfare for the new transitions. Qualitatively our results remain unchanged. Quantitatively, overall welfare is a little lower due to a higher level of expected medical costs (observe that C^* is lower than before, so the reason has to be a higher level of expected costs, rather than related to uncertainty). To compare to our original results, we we rescale the numbers in table 18 so that C^* equals that in the analysis above; after this

[28] We do not use these transition matrices in our main analysis out of a desire to have our transition matrices and risk aversion estimates correspond to the same population.

	Certainty Equivalent (!,000s)			Gains from LT contracts	
Income	C^*	CE_D	CE_{SPOT}	$\frac{C^*-CE_{SPOT}}{CE_{SPOT}}$	$\frac{CE_D-CE_{SPOT}}{C^*-CE_{SPOT}}$
Flat net	51.22	49.98	44.60	14.8%	81.3%
Non-mngr	51.22	41.01	38.14	34.3%	21.9%
Manager	81.55	48.72	47.37	72.2%	4.0%
Downs Mng	51.22	31.01	28.87	77.4%	9.6%

Table 18: Welfare for alternative transition matrices obtained for employees of a larger employer.

adjustment, all of the welfare numbers with the alternative transitions are within 3% of our basic numbers, reported in Section 5.3.[29]

6.4 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 would also not want to engage in savings in the first best. However, with spot contracting consumers may want to engage in precautionary savings to lower the costs of reclassification risk. Individuals can save in good states to weather periods of bad health. To study the impact of precautionary savings we solve a finite-horizon savings problem, with the same underlying fundamentals as in our main analysis, namely, the same income paths, 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.[30] 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").

[29] Note that we cannot compute an ACA welfare for this population because we lack data for this population that would allow us to determine full distributions of medical expense risk.

[30] 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 a expected present value of the 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.

	Certainty Equivalent (1,000s)			% Gains		
	(1)	(2)	(3)	(4)	(5)	(6)
Income	CE_D	CE_{SPOTwS}	CE_{SPOT}	$\frac{CE_D - CE_{SPOT}}{CE_{SPOT}}$	$\frac{CE_{SwS} - CE_{SPOT}}{CE_D - CE_{SPOT}}$	$\frac{CE_D - C_{SPOTwS}}{CE_{NSNB} - C_{SPOTwS}}$
Flat net	52.77	50.45	46.27	14.1%	64.3%	75.3%
Non-mngr	44.10	42.38	40.73	8.3%	49.0%	34.3%
Manager	51.77	50.82	50.32	2.9%	34.5%	18.1%
Downs mngr	34.10	32.75	31.74	7.4%	42.8%	26.1%

Table 19: Long-run welfare allowing for precautionary savings under spot contracts. The certainty equivalent of spot contracting with precautionary savings is denoted by CE_{SPOTwS}.

Table 19 shows the welfare effect of precautionary savings. As the spot contracting with precautionary savings outcome is feasible in our dynamic problem with one-sided commitment, CE_{SPOTwS} lies between CE_S and CE_D. Savings enable the consumer to transfer resources to future periods, to be consumed in periods of high marginal utility from consumption. Optimal dynamic contracts, however, do better, as they allow for state-specific savings. By charging state-contingent premiums the optimal contract enables the equation of consumption across all states in which the lapsation does not bind. In particular, the optimal contract transfers no resources to the most fortunate states.

Column (5) shows that precautionary savings closes between 42% and 65% of the gap between welfare with optimal dynamic contracts and spot contracting without savings. Column (6) gives the fraction of the welfare gap between the no-borrowing/no-savings constrained first-best outcome and the spot contracting with precuationary savings outcome that is closed by optimal dynamic contracts. As in our earlier analysis, dynamic contracts close a suignificant fraction of this gap with flat net income, but much less when net incomes are increasing with age.

6.5 Initial Health State

In our analysis we have assumed that a 25-year-old's health state is observable and priced when contracting starts. As a result, individuals who arrive at age 25 in different health states may face very different outcomes, and optimal dynamic contracts may be effective at insuring reclassification risk arising after age 25 to varying degrees depending on the age 25 health state.

Table **??** presents welfare when contracting starts in various health states at age 25, for various income paths. Starting in a good initial state has a limited impact on first best welfare (regardless of the income profile) because of the relatively temporary nature of a poor initial health state. However, it has a dramatic effect on the welfare achieved under

Welfare and Initial Health					
Initial health state	Non-manager welfare ($1,000s)				Required subsidy ($1,000s)
	C^*	CE_D	CE_{SPOT}	CE_{SPOTwS}	
Ex ante	53.67	44.10	40.73	42.19	
λ_1:					
1	53.76	47.20	41.79	43.70	–
2	53.56	46.05	40.82	42.71	4.20
3	53.35	44.55	39.36	41.21	9.10
4	53.29	44.34	39.23	41.05	10.27
5	53.08	42.10	37.37	39.25	15.62
6	52.86	39.47	37.26	37.87	21.07
7	52.50	33.16	33.04	33.10	30.43
$E[\cdot]$					1.93

Table 20: Non-manager welfare from alternative contracting institutions for alternative initial states λ_1, and the subsidies required to give all 25-year old non-manager consumers the same lifetime welfare with dynamic contracting as if they were in the healthiest state.

optimal dynamic contracts with one-sided commitment because starting in a poor health state makes front-loading much less attractive. This effect is greater the steeper the income profile. As a result, there is a dramatic effect of the age-25 health state of a consumer on her lifetime welfare under this regime. The situation with spot contracts (with or without precautionary savings) is intermediate: the effect of a poor age-25 health state is relative minor with a flat net income profile, but is greater with steeper income profiles because the high premium a consumer who is sick at age 25 incurs when young is costly because of her relatively low income.

These welfare results raise the question of how much ex ante welfare would be increased with optimal dynamic contracts if consumers were insured for the risk they bear arising from uncertain age-25 health, perhaps through some form of government insurance. The last column of Table **??** shows the subsidy that is required in each age-25 health state to give a consumer facing the dynamic contracting environment teh same welfare as if they were in the healthiest age-25 state. The expected value of these subsidies, using the probabilities of the various possible health states at age 25, is given in the last row of the table; the government would incur a cost of $1,930 per consumer to achieve this.

7 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. In contrast to the long-term contracts described in Proposition 1, they 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, no reclassification risk). In the context of our model, these policies provide a consumer who starts period 1 in the healthiest possible state ($\lambda_1 = 1$) with the guaranteed consumption path $\{y_t - p_t\}_{t=1}^T$, where the period t premium p_t is (we denote by $\lambda_t = 1$ the healthiest possible state in period t):[31]

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

To understand this formula, consider a case in which $T = 2$. In period 2 (the last period), the consumer pays a premium equal to $p_2 = \mathbb{E}[m_2|\lambda_2 = 1]$, the expected medical expenses of the healthiest possible period 2 consumer. In period 1, the consumer pays a premium of

$$\begin{aligned} p_1 &= \mathbb{E}[m_1|\lambda_1 = 1] - \delta\{\mathbb{E}[m_2|\lambda_1 = 1] - \mathbb{E}[m_2|\lambda_2 = 1]\} \\ &= \mathbb{E}[m_1|\lambda_1 = 1] - \delta \Pr(\lambda_2 > 1|\lambda_1 = 1)\{\mathbb{E}[m_2|\lambda_2 > 1] - \mathbb{E}[m_2|\lambda_2 = 1]\} \end{aligned}$$

The first term is the consumer's expected period 1 medical costs (since he starts with $\lambda_1 = 1$), 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 $\lambda_2 > 1$). In a problem in which $T = 3$, period 2 and 3 premiums would be as just described, while in the first period the consumer would prepay the value of the expected period 2 premium discount: $\Pr(\lambda_2 > 1|\lambda_1 = 1)\{\sum_{\tau=2,3} \delta^{\tau-1}\mathbb{E}[m_\tau|\lambda_2 > 1] - \mathbb{E}[m_\tau|\lambda_2 = 1]\} = \sum_{\tau=2,3} \delta^{\tau-1}\{\mathbb{E}[m_\tau|\lambda_1 = 1] - \mathbb{E}[m_\tau|\lambda_2 = 1]\}$.

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 path. The PKH contracts go to the extreme of completely preventing reclassification risk, resulting in a fully deterministic consumption path but excessively low

[31] We provide a proof in Appendix D.

initial consumption.[32] 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 offer optimal dynamic contracts: some of the healthiest types would actually be drawn away by a firm offering the contracts described in Proposition 1.[33]

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 next 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.[34] In each period t the consumer would pay a premium of

$$\mathbb{E}[m_t|\lambda_t] + \sum_{\tau>t} \delta^{\tau-t}\{\mathbb{E}[m_\tau|\lambda_t] - \mathbb{E}[m_\tau|\lambda_{t+1}=1]\} \tag{9}$$

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

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

equal to the change in expected medical expenses plus the repayment (with interest) of the second term in (9). Subtracting the period t payment [(10) modified to be for period t] from

[32] 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)]. In Section 5 we compare the utility generated by the optimal contracts of Proposition 1 with that arising from the PKH contracts.

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

[34] 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 ($\lambda_1 = 1$) and remains healthy ($\lambda_t = 1$ for all $t > 1$) would need to make payments in every period.

the period t premium (9), we see that the net premium payment in each period t is exactly the PKH premium (8).

7.1 Empirical Comparison of PKH and Optimal Dynamic Contracts

Using formula (8), we calculate that the initial PKH guaranteed renewable premium paid by a healthy 25-year old is \$5,736, about 27% higher than the initial premium paid by a healthy 25-year old individual in the optimal dynamic contract.[35] For a consumer with a with flat net income path, the excessively low iniital consumption required to eliminate all reclassification risk translates into a lower welfare: $CE_{PKH} = 49,509$, which is 3.6% lower than the certainty equivalent that this consumer would have with an optimal dynamic contract. The welfare loss from the PKH contract realtive to an optimal contract increases with rising income paths: for example, for a healthy 25-year old downscaled manager we find that $CE_{PKH} = 31,159$, resulting in a loss of 14.2% compared to an optimal dynamic contract.

8 Conclusion

In this paper we have studied the potential for long-term contracting in health insurance markets to insure against reclassification risk when insurers, but not consumers, can commit to the contract and consumers are unable to borrow to pre-pay premiums. We established the equilibrium properties of these contracts, and quantified their positive and normative implications relative to alternative market designs using detailed individual-level data on health risks, health transitions, insurance choices, and income profiles from the employees of a large firm. We compared these optimal dynamic contracts with one-sided commitment to (i) the first-best outcome (achieved by long-term contracts with two-sided commitment), (ii) spot contracts that are year-to-year, (iii) a benchmark in which all risk is insured but resources cannot be transferred across time, and (iv) ACA-like markets with spot contracts and community rating. We performed these comparisons for different income paths, including flat net income (where there is no motive for general consumption smoothing over time) and the rising income paths associated with both managers and non-managers at our firm.

Optimal dynamic contracts exhibit substantial front-loading at early ages, thereby enabling consumption guarantees that reduce reclassification risk. These guarantees can also

[35] In this section we use the transition matrices described in Section 6.3 for the employer with a larger population of covered individuals. The reason is that the PKH contract is very sensitive to the transition matrices later in life.

be provided by offering an equivalent guaranteed renewable premium path whose year-to-year increases match the consumer's income growth. Moreover, we show that when framed as guaranteed renewable contracts, these optimal contracts – which differ according to a consumer's income path – are self-selective, in the sense that consumers with differing income paths will all choose the contract that is designed for them.

Empirically, we find that optimal dynamic contracts for our population of consumers exhibit substantial front loading: a healthy consumer with a flat net income path at age 25, for example, pays a premium of $2,750 compared to expected medical costs of only $1,131. However, a healthy "downscaled manager," who income rises steeply over time would front load much less, paying a premium of $1,678. We find that the welfare implications of optimal dynamic contracts also depend crucially on the steepness of a consumer's income path. We find that optimal dynamic contracts with one-sided commitment would almost achieve the first-best for consumers with flat net income paths for whom front-loading is not costly. Conversely, dynamic contracts are only a little better than spot market contracts (with or without precautionary savings) for consumers with rising income profiles, for whom front-loading to facilitate reclassification risk insurance is quite costly. This intuition carries over to the welfare comparison between dynamic contracts and ACA-like markets: consumers with flatter income profiles prefer dynamic contracts but those with steeper income profiles prefer ACA-like markets.

We also consider several extensions of our main analysis. We find that lower consumer risk aversion, which reduces the costs of reclassification risk, makes optimal dynamic contracts nearly achieve the first best for all income paths. While lower risk aversion also improves the performance of other institutions, with lower risk aversion dynamic contracts come to be better than the ACA-like market for all income paths. (Of course, the differences in welfare are small when risk aversion is low.) Higher switching costs also improve the performance of dynamic contracts, which can achieve a level of welfare comparable to the ACA with switching costs on the order of $5,000. We also show that those consumers who are in poor health at age 25 do particularly poorly with dynamic contracts, as their high medical needs at that age preclude front loading. Insuring consumers against the risks of their age 25 health realization would cost the government approximately $2,000 per consumer.

In sum, dynamic contracts offer a way to substantially reduce reclassification risk without requiring community rating as in the ACA. However, with rising incomes we find that the ACA approach offers greater lifetime welfare, unless consumers' risk aversion is low (in which case reclassification risk is not much of a problem) or consumers face significant switching costs.

References

[1] Abramitzky, R.(2008) "The Limits of Equality: Evidence from the Israeli Kibbutz," *Quarterly Journal of Economics* 123: 1111-1159.

[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] 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.

[5] 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.

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

[7] 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.

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

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

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

[11] Finkelstein, A, K. McGarry and A. Sufi. (2005) "Dynamic Inefficiencies in Insurance Markets: Evidence from Long-Term Care Insurance" *The American Economic Review, Papers and Proceedings*, 95: 224-228.

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

[13] 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.

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

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

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

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

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

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

[20] 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.

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

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

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

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

[25] 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.

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

In this appendix, we characterize equilibrium contracts with one-sided commitment, allowing for a switching cost $\sigma \geq 0$. Proposition 1 follows as the special case in which $\sigma = 0$. Specifically, we establish the following result:

Proposition 3 *The equilibrium contract in a competitive market with one-sided commitment when the consumer cannot borrow and has a switching cost $\sigma \geq 0$ is characterized by the consumption levels offered in the first period of a switching contract starting in period t with health state λ_t, $c_t^\sigma(\lambda_t)$, and the consumption offers in period 1, $c_1(\lambda_1)$. In the equilibrium contract, the consumer is fully insured against within-period medical expense risk, and enjoys in each period t following health state history $(\lambda_1, ..., \lambda_t)$ the certain consumption $\max\{c_1(\lambda_1), \max_{\tau \in [2,t]} c_\tau^\sigma(\lambda_\tau)\}$. The levels $c_t^\sigma(\lambda_t)$ lead a firm signing a switching consumer to break even in expectation.*[36] *Consumers have no incentive to save given the equilibrium contract.*

We begin by specifying the contract design problem recursively, and then provide a series of characterization results that collectively establish Proposition 3.

9.1 The period T problem

Suppose that the period T state is $(\lambda_T, S_T, \sigma_T)$, where S_T is the discounted expected loss that can be sustained from that period on, and σ_T is the switching cost the consumer incurs in period T (which equals 0 if he is not switching and equals σ if he is). Consider the following problem, whose value is $\overline{V}_T(\lambda_T, S_T, \sigma_T)$:

$$\begin{array}{cc} \max_{c_T(\cdot)} & \int u(c_T(m_T) - \sigma_T) dF_T(m_T|\lambda_T) \\ s.t. & \int c_T(m_T) dF_T(m_T|\lambda_T) \leq S_T + y_T - \mathbb{E}[m_T|\lambda_T];\ \gamma_T \end{array} \tag{11}$$

The FOC for this problem is

$$u'(c_T(m_T) - \sigma_T) = \gamma_T \text{ for all } m_T$$

So we have

$$c_T(m_T) = S_T + y_T - \mathbb{E}[m_T|\lambda_T]$$

[36] Equivalently, this is the consumption guarantee that would be offered to a consumer who incurs no switching costs by firms that must earn an expected discounted profit equal to σ.

and

$$\overline{V}_T(\lambda_T, S_T, \sigma_T) = u(S_T + y_T - \mathbb{E}[m_T|\lambda_T] - \sigma_T)$$

Observe, that we can rewrite problem (11) as

$$\begin{array}{ll} \max_{C_T(\cdot)} & \int u(C_T(m_T))dF_T(m_T|\lambda_T) \\ s.t. & \int (C_T(m_T) + \sigma_T)dF_T(m_T|\lambda_T) \leq S_T + y_T - \mathbb{E}[m_T|\lambda_T];\ \gamma_T \end{array}$$

and in turn as

$$\begin{array}{ll} \max_{C_T(\cdot)} & \int u(C_T(m_T))dF_T(m_T|\lambda_T) \\ s.t. & \int C_T(m_T)dF_T(m_T|\lambda_T) \leq (S_T - \sigma_T) + y_T - \mathbb{E}[m_T|\lambda_T];\ \gamma_T \end{array} \tag{12}$$

which shows that $\overline{V}_T(\lambda_T, S_T, \sigma_T) = \overline{V}_T(\lambda_T, S_T - \sigma_T, 0)$. That is, having the consumer incur a consumption loss of σ_T due to switching is equivalent to having the loss that can be sustained by the firm be $S_T - \sigma_T$. Intuitively, we can think of the firm as first compensating the consumer for the switching cost, and then devising the optimal consumption level given a savings budget of $S_T - \sigma_T$.

It will be convenient to define the value function $V_T(\lambda_T, S_T - \sigma_T) \equiv \overline{V}_T(\lambda_T, S_T - \sigma_T, 0)$, which is the value of problem (12). Observe that this value function is strictly concave in S_T since $\partial^2 V_T(\lambda_T, S_T)/\partial S_T^2 = u''(S_T + y_T - \mathbb{E}[m_T|\lambda_T])$, and by Assumption A1 we also have strict increasing differences in (λ_T, S_T) since $\partial V_T(\lambda_T, S_T)/\partial S_T = u'(S_T + y_T - \mathbb{E}[m_T|\lambda_T])$ is strictly increasing in λ_T.

9.2 The period t problem

Suppose the period t state is $(\lambda_t, S_t, \sigma_t)$. (As above, σ_t equals 0 if the consumer is not switching and equals σ if he is.) Consider the following problem, whose value is $\overline{V}_t(\lambda_t, S_t, \sigma_t)$:

$$\begin{array}{ll} \max_{c_t(\cdot), S_{t+1}(\cdot)} & \int [u(c_t(m_t, \lambda_{t+1}) - \sigma_t) + \delta_t \overline{V}_{t+1}(\lambda_{t+1}, S_{t+1}(m_t, \lambda_{t+1}), 0)]dF_t(m_t, \lambda_{t+1}|\lambda_t) \\ s.t. & \int [c_t(m_t, \lambda_{t+1}) + \delta_t S_{t+1}(m_t, \lambda_{t+1})]dF_t(m_t, \lambda_{t+1}|\lambda_t) \leq S_t + y_t - \mathbb{E}[m_t|\lambda_t];\ \gamma_t \\ & \overline{V}_{t+1}(\lambda_{t+1}, S_{t+1}(m_t, \lambda_{t+1}), 0) \geq \overline{V}_{t+1}(\lambda_{t+1}, 0, \sigma) \text{ for all } (m_t, \lambda_{t+1});\ \mu_t(m_t, \lambda_{t+1}) \end{array} \tag{13}$$

Remark 1 *The second constraint is the lapsation constraint.*

Remark 2 *Note that $c_t(m_t, \lambda_{t+1})$ and $S_{t+1}(m_t, \lambda_{t+1})$ also depend on $(\lambda_t, S_t, \sigma_t)$, the state entering period t. We do not include these arguments to simplify notation. Below, for example, we report some results on how changes in the state affect $c_t(m_t, \lambda_{t+1})$ and $S_{t+1}(m_t, \lambda_{t+1})$.*

Following similar logic as for the period T problem, we can rewrite problem (13) as

$$\begin{aligned} \max_{C_t(\cdot), S_{t+1}(\cdot)} \quad & \int [u(C_t(m_t, \lambda_{t+1})) + \delta_t \overline{V}_{t+1}(\lambda_{t+1}, S_{t+1}(m_t, \lambda_{t+1}), 0)] dF_t(m_t, \lambda_{t+1}|\lambda_t) \\ s.t. \quad & \int [C_t(m_t, \lambda_{t+1}) + \delta_t S_{t+1}(m_t, \lambda_{t+1})] dF_t(m_t, \lambda_{t+1}|\lambda_t) \leq (S_t - \sigma_t) + y_t - \mathbb{E}[m_t|\lambda_t]; \ \gamma_t \\ & \overline{V}_{t+1}(\lambda_{t+1}, S_{t+1}(m_t, \lambda_{t+1}), 0) \geq \overline{V}_{t+1}(\lambda_{t+1}, 0, \sigma) \text{ for all } (m_t, \lambda_{t+1}); \ \mu_t(m_t, \lambda_{t+1}) \end{aligned} \tag{14}$$

Thus, we have:

Lemma 1 *$\overline{V}_t(\lambda_t, S_t, \sigma_t) = \overline{V}_t(\lambda_t, S_t - \sigma_t, 0)$ for all $(\lambda_t, S_t, \sigma_t)$*

As in the period T problem, it is convenient to define the value function $V_t(\lambda_t, S_t - \sigma_t) \equiv \overline{V}_t(\lambda_t, S_t - \sigma_t, 0)$

Corollary 1 *The lapsation constraint can be written as $S_t(m_t, \lambda_{t+1}) \geq -\sigma$.*

Proof. From Lemma 1, the inequality $\overline{V}_{t+1}(\lambda_{t+1}, S_{t+1}, 0) \geq \overline{V}_{t+1}(\lambda_{t+1}, 0, \sigma)$ is equivalent to the inequality $\overline{V}_{t+1}(\lambda_{t+1}, S_{t+1}, 0) \geq \overline{V}_{t+1}(\lambda_{t+1}, -\sigma, 0)$, which holds if and only if $S_t \geq -\sigma$.

■

As a result, we can rewrite problem (14) as the following problem, whose value is $V_t(\lambda_t, S_t - \sigma_t)$:

$$\begin{aligned} \max_{c_t(\cdot), S_{t+1}(\cdot)} \quad & \int [u(c_t(m_t, \lambda_{t+1})) + \delta_t V_{t+1}(\lambda_{t+1}, S_{t+1}(m_t, \lambda_{t+1}))] dF_t(m_t, \lambda_{t+1}|\lambda_t) \\ s.t. \quad & \int [c_t(m_t, \lambda_{t+1}) + \delta_t S_{t+1}(m_t, \lambda_{t+1})] dF_t(m_t, \lambda_{t+1}|\lambda_t) \leq S_t - \sigma_t + y_t - \mathbb{E}[m_t|\lambda_t]; \ \gamma_t \\ & \delta_t f_t(m_t, \lambda_{t+1}|\lambda_t) \cdot S_{t+1}(m_t, \lambda_{t+1}) \geq -\delta_t f_t(m_t, \lambda_{t+1}|\lambda_t) \cdot \sigma \text{ for all } (m_t, \lambda_{t+1}); \ \mu_t(m_t, \lambda_{t+1}) \end{aligned} \tag{15}$$

Remark 3 *We write the second constraint with $\delta_t f_t(m_t, \lambda_{t+1}|\lambda_t)$ (which is strictly positive) to simplify statement of the FOC (17) below.*

9.3 Equilibrium Contract Characterization

The equilibrium contract solves (12) and (15) with $\sigma_t = 0$ for all t. We now characterize its solution. We do so through an inductive argument. We have already seen that $V_T(\lambda_T, S_T)$ is strictly concave in S_T and has strictly increasing differences in (λ_T, S_T). In considering the period t problem, we now assume that $V_{t+1}(\lambda_{t+1}, S_{T+1})$ is strictly concave in S_{t+1} and has strictly increasing differences in (λ_{t+1}, S_{t+1}). The characterization results below will imply that the same properties hold for the value function $V_t(\lambda_t, S_t)$.

The FOC for problem (15) are:

$$\begin{aligned} u'(c_t(m_t, \lambda_{t+1})) &= \gamma_t \text{ for all } (m_t, \lambda_{t+1}) && (16) \\ \frac{\partial V_{t+1}(\lambda_{t+1}, S_{t+1}(m_t, \lambda_{t+1}))}{\partial S_{t+1}} &= \gamma_t - \mu_t(m_t, \lambda_{t+1}) \text{ for all } (m_t, \lambda_{t+1}) && (17) \end{aligned}$$

Condition (16) implies that c_t is independent of (m_t, λ_{t+1}). We can therefore rewrite problem (15) for $\sigma_t = 0$, whose value is $V_t(\lambda_t, S_t)$, as:

$$\begin{aligned} \max_{c_t, S_{t+1}(\cdot)} \quad & u(c_t) + \delta_t \int V_{t+1}(\lambda_{t+1}, S_{t+1}(m_t, \lambda_{t+1}))dF_t(m_t, \lambda_{t+1}|\lambda_t) \\ s.t. \quad & c_t + \delta_t \int S_{t+1}(m_t, \lambda_{t+1})]dF_t(m_t, \lambda_{t+1}|\lambda_t) \leq S_t + y_t - \mathbb{E}[m_t|\lambda_t];\ \gamma_t \\ & \delta_t f_t(m_t, \lambda_{t+1}|\lambda_t) \cdot S_{t+1}(m_t, \lambda_{t+1}) \geq -\delta_t f_t(m_t, \lambda_{t+1}|\lambda_t) \cdot \sigma \text{ for all } (m_t, \lambda_{t+1});\ \mu_t(m_t, \lambda_{t+1}) \end{aligned} \quad (18)$$

The FOC for this problem are (17) and

$$u'(c_t) = \gamma_t \text{ for all } (m_t, \lambda_{t+1}) \quad (19)$$

We now establish a number of lemmas characterizing the solution of this problem.

Lemma 2 *$V_t(\lambda_t, S_t)$ is strictly concave in S_t.*

Proof. Let $[c_t, S_{t+1}(\cdot)]$ be the solution for (λ_t, S_t) and $[c'_t, S'_{t+1}(\cdot)] \neq [c_t, S_{t+1}(\cdot)]$ be the solution of problem (18) for (λ_t, S'_t). Consider the problem for (λ_t, S^ρ_t) where $S^\rho_t = \rho S_t + (1-\rho)S'_t$ for $\rho \in (0,1)$. In this problem $[c^\rho_t, S^\rho_{t+1}(\cdot)]$ is feasible, where $c^\rho_t = \rho c_t + (1-\rho)c'_t$ and $S^\rho_{t+1} = \rho S_{t+1} + (1-\rho)S'_{t+1}$. Moreover, because $u(\cdot)$ is strictly concave and $V_{t+1}(\lambda_{t+1}, S_{t+1})$ is strictly concave in S_{t+1}, $[c^\rho_t, S^\rho_{t+1}(\cdot)]$ achieves a value strictly larger than $\rho V_t(\lambda_t, S_t) + (1-\rho)V_t(\lambda_t, S'_t)$, which implies that $V_t(\lambda_t, S^\rho_t) > \rho V_t(\lambda_t, S_t) + (1-\rho)V_t(\lambda_t, S'_t)$. ■

Lemma 3 *c_t is strictly increasing in S_t, and γ_t is strictly decreasing in S_t.*

Proof. By the Envelope Theorem, $\partial V_t(\lambda_t, S_t)/\partial S_t = \gamma_t$. Since $V_t(\lambda_t, S_t)$ is strictly concave in S_t, γ_t is strictly decreasing in S_t. By (19), this implies that c_t is strictly increasing in S_t. ■

Lemma 4 *$S_{t+1}(\cdot)$ is independent of m_t.*

Proof. Suppose $S_{t+1}(m_t, \lambda_{t+1}) > S_{t+1}(m_t', \lambda_{t+1})$ for some m_t and $m_t' \neq m_t$. Then

$$\frac{\partial V_{t+1}(\lambda_{t+1}, S_{t+1}(m_t', \lambda_{t+1}))}{\partial S_{t+1}} > \frac{\partial V_{t+1}(\lambda_{t+1}, S_{t+1}(m_t, \lambda_{t+1}))}{\partial S_{t+1}}$$

Moreover, since $S_{t+1}(m_t, \lambda_{t+1}) > S_{t+1}(m_t', \lambda_{t+1}) \geq -\sigma$, we have $\mu_t(m_t', \lambda_{t+1}) \geq \mu_t(m_t, \lambda_{t+1}) = 0$. But this contradicts (17). ■

So, letting $\overline{f}_t(\lambda_{t+1}|\lambda_t)$ be the marginal conditional distribution of λ_{t+1} given λ_t, we can rewrite the period t problem (for $\sigma_t = 0$) as

$$\begin{aligned} \max_{c_t, S_{t+1}(\cdot)} \quad & u(c_t) + \delta_t \textstyle\int V_{t+1}(\lambda_{t+1}, S_{t+1}(\lambda_{t+1}))\overline{f}_t(\lambda_{t+1}|\lambda_t)d\lambda_{t+1} \\ s.t. \quad & c_t + \delta_t \textstyle\int S_{t+1}(\lambda_{t+1})\overline{f}_t(\lambda_{t+1}|\lambda_t)d\lambda_{t+1} \leq S_t + y_t - \mathbb{E}[m_t|\lambda_t];\ \gamma_t \\ & \delta_t \overline{f}_t(\lambda_{t+1}|\lambda_t) \cdot S_{t+1}(\lambda_{t+1}) \geq -\delta_t \overline{f}_t(\lambda_{t+1}|\lambda_t) \cdot \sigma \text{ for all } \lambda_{t+1};\ \mu_t(\lambda_{t+1}) \end{aligned} \tag{20}$$

Its FOC are (19) and

$$\frac{\partial V_{t+1}(\lambda_{t+1}, S_{t+1}(\lambda_{t+1}))}{\partial S_{t+1}} = \gamma_t \ - \mu_t(\lambda_{t+1}) \ \text{ for all } (\lambda_{t+1}) \tag{21}$$

Lemma 5 *For all λ_t, $S_{t+1}(\lambda_{t+1})$ is nondecreasing in S_t at all λ_{t+1}.*

Proof. Consider $S_t > S_t'$. Denote the solutions $[c_t, S_{t+1}(\cdot)]$ and $[c_t', S_{t+1}'(\cdot)]$ respectively. Consider state λ_{t+1}. The result is immediate if $S_{t+1}'(\lambda_{t+1}) = -\sigma$ so suppose that $S_{t+1}'(\lambda_{t+1}) > -\sigma$. Then $\mu_t'(\lambda_{t+1}) = 0$. Since $\gamma_t < \gamma_t'$ by Lemma 3 and $\mu_t(\lambda_{t+1}) \geq \mu_t'(\lambda_{t+1})$, (21) implies that we must have

$$\frac{\partial V_{t+1}(\lambda_{t+1}, S_{t+1}'(\lambda_{t+1}))}{\partial S_{t+1}} > \frac{\partial V_{t+1}(\lambda_{t+1}, S_{t+1}(\lambda_{t+1}))}{\partial S_{t+1}}$$

which, in turn, implies that $S_{t+1}(\lambda_{t+1}) > S_{t+1}'(\lambda_{t+1})$ by the strict concavity of $V_{t+1}(\lambda_{t+1}, S_{t+1})$ in S_{t+1}. ■

Lemma 6 *$S_{t+1}(\cdot)$ is nondecreasing in λ_{t+1}.*

Proof. Suppose $S_{t+1}(\lambda_{t+1}) > S_{t+1}(\lambda'_{t+1})$ where $\lambda_{t+1} < \lambda'_{t+1}$. Then

$$\frac{\partial V_{t+1}(\lambda'_{t+1}, S_{t+1}(\lambda'_{t+1}))}{\partial S_{t+1}} > \frac{\partial V_{t+1}(\lambda_{t+1}, S_{t+1}(\lambda'_{t+1}))}{\partial S_{t+1}} > \frac{\partial V_{t+1}(\lambda_{t+1}, S_{t+1}(\lambda_{t+1}))}{\partial S_{t+1}},$$

where the first inequality follows because $V_{t+1}(\lambda_{t+1}, S_{t+1})$ has strict increasing differences in (λ_{t+1}, S_{t+1}) and the second because $V_{t+1}(\lambda_{t+1}, S_{t+1})$ is strictly concave in S_{t+1}. Moreover, since $S_{t+1}(\lambda_{t+1}) > S_{t+1}(\lambda'_{t+1}) \geq -\sigma$, we have $\mu_t(\lambda'_{t+1}) \geq \mu_t(\lambda_{t+1}) = 0$. But this contradicts (21). ■

Lemma 7 *γ_t is strictly increasing in λ_t.*

Proof. Suppose that $\lambda'_t > \lambda_t$ but $\gamma'_t \leq \gamma_t$. Denote the solutions for λ_t and λ'_t by $[c_t, S_{t+1}(\cdot)]$ and $[c'_t, S'_{t+1}(\cdot)]$ respectively. Then, by (21), for all λ_{t+1} such that $S_{t+1}(\lambda_{t+1}) > -\sigma$, we have $S'_{t+1}(\lambda_{t+1}) \geq S_{t+1}(\lambda_{t+1})$. Thus, $S'_{t+1}(\lambda_{t+1}) \geq S_{t+1}(\lambda_{t+1})$ at all λ_{t+1}. Assumptions A1 and A2, plus Lemma 6, then imply that

$$\begin{aligned} & S_t + y_t - \mathbb{E}[m_t|\lambda_t] - \delta_t - \int S_{t+1}(\lambda_{t+1}) \overline{f}_t(\lambda_{t+1}|\lambda_t) d\lambda_{t+1} \\ \geq \; & S_t + y_t - \mathbb{E}[m_t|\lambda'_t] - \delta_t - \int S_{t+1}(\lambda_{t+1}) \overline{f}_t(\lambda_{t+1}|\lambda'_t) d\lambda_{t+1} \\ \geq \; & S_t + y_t - \mathbb{E}[m_t|\lambda'_t] - \delta_t - \int S'_{t+1}(\lambda_{t+1}) \overline{f}_t(\lambda_{t+1}|\lambda'_t) d\lambda_{t+1} \end{aligned}$$

so the budget constraint implies that $c'_t < c_t$. But this violates (19) when $\gamma'_t \leq \gamma_t$ – a contradiction. ■

Lemma 8 *$V_t(\lambda_t, S_t)$ has strictly increasing differences in (λ_t, S_t).*

Proof. Since, by the Envelope Theorem, $\partial V_t(\lambda_t, S_t)/\partial S_t = \gamma_t$, the result follows from Lemma 7. ■

Lemmas 2 and 8 establish that the induction hypothesis is valid. We now show that the above lemmas establish Proposition 3.

Note from (21) for the period t problem that the states in which the lapsation constraint does not bind (so that $S_{t+1} > -\sigma$) have equal marginal utilities of S_{t+1}, equal to γ_t. This implies (using the Envelope Theorem for the period t+1 problems) that they have equal γ_{t+1}, which in turn implies that they have equal c_{t+1}. Moreover, $\gamma_t = \gamma_{t+1}$ in those states, so $c_t = c_{t+1}$. The states in which lapsation constraint does bind in the period t problem have marginal utilities of S_{t+1} that are strictly less than γ_t. Hence, again from the Envelope

Theorem, in these states we have $\gamma_{t+1} < \gamma_t$, which implies that $c_{t+1} > c_t$. Thus, the contract can be thought of as starting with a guaranteed consumption, which gets bumped up to a new higher guaranteed consumption whenever the lapsation constraint binds. The new higher guaranteed consumption level is exactly the first-period consumption in the optimal contract that would start at that date and health state if the consumer were to start with a new firm, incurring the switching cost σ (since $S_{t+1} = -\sigma$ at that continuation).[37]

Finally, observe as well that γ_t is weakly falling (and c_t is weakly increasing) along any realized path. As a result, the consumer never has any desire to save.

10 Appendix B: Self Selection

We first establish the following Lemma:

Lemma 9 *Let $p = (p_1, ..., p_T)$ and $\widehat{p} = (\widehat{p}_1, ..., \widehat{p}_T)$ be guaranteed premium paths such that $\widehat{p} \geq p$. Suppose that (i) the insurer earns a non-negative expected NPV when guaranteed premium path p is chosen by a y-type consumer in health state λ_1 (given the consumer's optimal lapsation behavior), and (ii) in every period $t > 1$, if a y-type consumer in health state λ_t optimally does not lapse when facing path p, then the insurer's expected continuation profit is non-positive. Then the insurer's expected NPV is non-negative when premium path $\widehat{p}$ is chosen by a y-type consumer in health state λ_1.*

Proof. Let $U(t, \lambda_t)$ and $\widehat{U}(t, \lambda_t)$ denote the insured's continuation payoff at state (t, λ_t) given optimal lapsation behavior under p and $\widehat{p}$ respectively. Let $\Lambda(t)$ and $\widehat{\Lambda}(t)$ denote the set of health states that optimally lapse in period t, under p and $\widehat{p}$ respectively; $\Lambda_{NL}(t)$ and $\widehat{\Lambda}_{NL}(t)$ are the complemetary sets who do not lapse. Finally, let $\Pi(t, \lambda_t)$ and $\widehat{\Pi}(t, \lambda_t)$ denote the insured's contiuation payoff at state (t, λ_t) given optimal lapsation behavior under p and $\widehat{p}$ respectively. Assumption (i) therefore says that $\Pi(1, \lambda_1) \geq 0$, while assumption (ii) says that $\Pi(t, \lambda_t) \leq 0$ for all $t > 1$ and $\lambda_t \in \Lambda_{NL}(t)$ [of course, $\Pi(t, \lambda_t) = 0$ for all $\lambda_t \in \Lambda(t)$].

Note, first, that $U(t, \lambda_t) \geq \widehat{U}(t, \lambda_t)$: starting in period t, the insured who faces p could adopt the same lapsation behavior as when facing $\widehat{p}$ and receive a weakly higher payoff since he would be facing lower prices. His optimal lapsation behavior yields a still higher payoff. Next, the fact that $U(t, \lambda_t) \geq \widehat{U}(t, \lambda_t)$ implies that $\Lambda(t) \subseteq \widehat{\Lambda}(t)$: any type who lapses in period t when facing p also lapses when facing $\widehat{p}$. Finally, consider the expected profit of the firm starting at state $(1, \lambda_1)$. This is the probability weighted average of the payoffs achieved along the various possible sequences of health states $(\lambda_1, ..., \lambda_T)$. For each

[37] Equivalently, it is the guaranteed consumption level that would be offered to a consumer who incurs no switching costs but faces firms who must earn an expected discounted profit equal to σ.

each sequence the firm earns premiums and incurs costs until the insured lapses. Since $\Lambda(t) \subseteq \widehat{\Lambda}(t)$, each such sequence hits lapsation weakly earlier under $\widehat{p}$ than under p. Since $\Pi(t, \lambda_t) \leq 0$ for all $t > 1$ and $\lambda_t \in \Lambda_{NL}(t)$, the earlier termination behavior under $\widehat{p}$ (but earning the same premiums p) – weakly raises the expected profit earned for the sequence, by changing a nonpositive expected continuation 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, further raises the expected profits earned from this health state sequence. As a result, $\Pi(1, \lambda_1) \leq \widehat{\Pi}(1, \lambda_1)$. ■

We next establish the following Lemma:

Lemma 10 *The insurer earns a nonnegative expected NPV if type y' in health state λ_1 chooses the guranteed premium path p_y that is intended for a type y consumer in health state λ_1.*

Proof. The proof is by induction on both the number of periods T and the number of periods that remain. First, observe that the result is clearly true if $T = 1$ as lapsation behavior is then not an issue and expected medical expenses don't depend on the consumer's income level. So, assume that the result is true for all $T < \widehat{T}$. We will show that it is true for $T = \widehat{T}$. To do so, we use another induction argument.

Assume the following Induction Hypothesis: If starting in period $t \in [2, \widehat{T}]$ the consumer has not yet lapsed, lapsation behavior of type $y\prime$ starting in period t is either (A) the same as for type y (meaning, after any previous history of health states and sequence of decisions not to lapse), or (B) different and raises the expected continuation NPV of the insurer starting in period t.

Observe that the Induction Hypothesis holds if $t = \widehat{T}$, since then lapsation behavior is the same for type $y\prime$ as for type y – both types lapse iff $E[m_{\widehat{T}}|\lambda_{\widehat{T}}] < p_{\widehat{T}}$.

Now suppose that the Induction Hypothesis holds for periods $t, ..., \widehat{T}$, where $t \geq 3$, and consider period $t - 1$ (≥ 2) after some previous history of health states and a sequence of decisions in which the insured has not yet lapsed. Suppose first that lapsation behavior is different for type $y\prime$ than for type y in a health state λ_{t-1} in which type y would not lapse. Then since $S_{t-1}(\lambda_{t-1}) \geq 0$ in the solution to the type y problem in period $t - 1$ (recall that S_{t-1} may also depend on the previous history $\lambda_1, ..., \lambda_{t-2}$), lapsation by type $y\prime$ in state λ_{t-1} would remove a continuation that had a weakly negative continuation expected NPV for the insurer and replace it with a zero NPV.

Suppose, instead, that state λ_{t-1} is one in which type y would lapse in period $t - 1$, but instead type $y\prime$ does not lapse. We will show that this changes what would have been a zero NPV continuation into a continuation with a nonnegative NPV. Let p'' be the premium path offered to a known y-type in health state λ_{t-1} who lapses. When a y-type chooses this

premium path the insurer breaks even (by the competitive market assumption). Note that $p''_\tau = p_{y\tau} - \Delta$ for $\tau \geq t-1$ and $\Delta > 0$ (since the two guaranteed premium paths differ only in offering different initial premium caps and then the premium change each period equals the change in the y-type's income.

We first argue that if a y'-type in health state λ_{t-1} were to choose premium path p'' the insurer would earn a nonnegative expected NPV. This follows since this result (i.e., Lemma 10) holds for contract periods of length $T < \widehat{T}$ (which is the case for contracts offered to lapsing consumers after period 1). But now, invoking Lemma 9 and the fact that $p''_\tau = p_{y\tau} - \Delta$ for $\tau \geq t-1$ and $\Delta > 0$, the insurer's expected NPV when the y'-type in health state λ_{t-1} does not lapse (and, hence, stick with p_y) in period $t-1$ must be nonnegative. So applying induction, properties (A) and (B) hold for all periods $t = 1, ..., \widehat{T}$.

Finally, consider period 1. If a y'-type in health state λ_1 chooses the premium path p_y intended for a type y consumer in health state λ_1, the insurer's first period costs are the same as if a y-type in health state λ_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 2 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 NPV must be weakly higher. Thus, the insurer must have a nonnegative expected NPV when a y'-type chooses contract p_y. ■

We now prove Proposition 2:

Proof. of Proposition 2: We argue that the menu is self-selective. From Lemma 10 we know that if a y-type chooses the plan intended for it the insurer breaks even but if it instead chooses the policy intended for type y' the insurer earns a non-negative profit. But the policy intended for the y-type maximizes the y-type's discounted expected utility subject to the constraint that the insurer at least break-even. The policy intended for type y' was therefore feasible, which implies that it cannot be preferred by type y. ■

11 Appendix C: Effects of Health State Persistence

In this Appendix we explore the effects of health state persistence on the welfare achieved under various contracting institutions. To illustrate the effects of greater persistence, we derive the optimal contracts and corresponding welfare using transition matrices representing the weighted average of the identity matrix (with weight $\rho \in \{0, 1]$) and uniform transition probabilities (namely, from any state there is 1/7 probability to transition to any state). The identity matrix represents full persistence, which is captured by $\rho = 1$. Instead for low ρ the states become less persistent, and in the extreme case of $\rho = 0$ the state is independent

ρ	C^*	C_{SPOT}	CE_D	C_{SwS}	$\frac{CE_D}{CE_{SPOT}}$	$\frac{CE_D}{CE_{SPOTwS}}$
0	47.05	39.83	45.59	44.45	1.14	1.03
0.3000	47.05	39.83	45.17	43.45	1.13	1.04
0.5000	47.05	39.83	44.72	42.68	1.12	1.05
0.7000	47.05	39.83	43.94	41.78	1.10	1.05
0.9000	47.05	39.83	42.12	40.64	1.06	1.04
0.9999	47.05	39.83	39.84	39.83	1.00	1.00

Table 21: Welfare for alternative transition matrices capturing health state persistence. With probability ρ the state is unchanged, and with probability $1-\rho$ the state is a random draw, with equal probabilities, of the seven possible states.

of past health. Note also that, with this transition process, expected medical expenses are independent of ρ. We assume here flat income (since expected health care costs are independent of age, this also represents flat net income).

Table 21 reports the welfare achieved in the first best (C^*), with spot contracting (CE_{SPOT}), with optimal dynamic contracts with one-sided commitment (C_D), and with spot contracts with precuationary savings possible (CE_{SPOTwS}).[38] (Since we have flat net income, $C^*_{NBNS} = C^*$ here.) The last two columns show the percentage welfare gain from optimal dynamic contracts versus spot contracting and versus spot contracting with precautionary savings. Observe that greater persistence has no effect on welfare for either the first best or spot contracting regimes. The first best is unaffected by persistence because expected medical expenses are independent of ρ. Welfare with spot contracting is also independent of ρ because the ex ante distribution of premiums in any period is independent of ρ and preferences are additive across time periods. Once precautionary savings is allowed, however, welfare is greater the lower is ρ, as precautionary savings can smooth out short-run shocks to medical expenses, but accomplishes less when those shocks are persistent; as shocks become perfectly persistent, precautionary savings provides no benefit at all. Finally, the welfare achievable with optimal dynamic contracts is declining in the level of persistence. Indeed, in the limit as ρ approaches 1 welfare with dynamic contracts approaches the level with spot contracting (with or without precautionary savings) as, in that limiting case, there is no reclassification risk to insure once the consumer's age-25 health state is realized. The gains of optimal contracts over spot contracts declines montonically in ρ, while it is inverse U-shaped when there is precautionary savings.

[38] Precuautionary savings is introduced in Section 6.4.

12 Appendix D: 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 $p_{t+1}(\lambda_{t+1}) - p_{t+1}(1)$ [where $p_{t+1}(1)$ is the period $t+1$ premium for the healtiest period $t+1$ health state, $\lambda_{t+1} = 1$, 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 healtiest type would pay.

The premiums for these contracts will in equilibrium be:

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

and for $t < T$,

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

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

Lemma 11 *For all t, $p_t(\lambda_t) = \mathbb{E}[m_t|\lambda_t] + \sum_{\tau>t} \delta^{\tau-t}\{\mathbb{E}[m_\tau(\lambda_\tau)|\lambda_t] - \mathbb{E}[m_\tau(\lambda_\tau)|\lambda_{t+1} = 1]\}$*

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:

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

■