

The Welfare Effects of Long-Term Health Insurance Contracts*

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

January 2017
(This Draft: May 2017)

Abstract

Reclassification risk is a major concern in health insurance. We use a rich dataset with individual-level information on health risk to empirically study one possible solution: dynamic contracts. Empirically, dynamic contracts with one-sided commitment substantially reduce the reclassification risk present with spot contracting, achieving close to the first-best for consumers with flat net income paths. Gains are smaller for consumers with net income growth, and these consumers prefer ACA-like community rating over dynamic contracts. However, lower risk aversion, sufficient switching costs, or government insurance of pre-age-25 health risks can raise welfare with dynamic contracts above the level in ACA-like markets.

1 Introduction

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

* We thank Neale Mahoney and Nathan Hendren, as well as seminar participants at Arizona, Bocconi, EIEF, Harvard, the MIT Public Finance lunch, the April 2017 NBER joint public finance/insurance conference, Penn, Princeton, Stanford, Toulouse, UCLA, UNC, and Washington University for their comments. All authors are grateful for support from NSF grant SES-1259770. 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

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, and we focus on this 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.

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

¹See, for example, the discussions in Diamond (1992), Cochrane (1995), and Pauly, Kunreuther, and Hirth (1995). Another reason the first best may be unattainable, which we do not model, is moral hazard.

²Were consumers able to borrow freely, the first best could be achieved by having consumers pay all premiums up front. This financial imperfection likely stems from similar factors as those that prevent committing consumers to make large ex post payments to an insurer.

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 can be equivalently offered as guaranteed premium path contracts, and when offered in this form are self-selective, in the sense that consumers with different lifetime income paths and risk preferences prefer the contracts designed for them.³

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 that 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 state transition process that captures how this risk evolves over time. We use as well 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: front-loading is highest for individuals in medium health states. 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

³Guaranteed premium path contracts contractually specify the path of future premiums at which the consumer can continue coverage if she has not previously lapsed. These are distinguished by their premium guarantees from what are called "guaranteed renewable" contracts. The latter merely state that the consumer has a right to renew at a rate at the insurer's discretion, but that must be the same as what the insurer offers to all other consumers in the same policy. This discretion potentially allows the insurer to induce lapsation by raising premiums or degrading quality (effectively cancelling the policy) and to then re-enroll consumers in new policies at rates based on their health states. In practice, in the pre-ACA world, an insurer's ability to do this varied across states due to differences in state regulatory stringency.

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 the 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: a downscaled 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 and provide less insurance, due to the lower degree of front-loading.

We then consider the welfare effects of these long-term contracts. 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. These contracts offer full insurance (generating a deterministic consumption path), 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. Still, this is a very large loss in lifetime expected utility.

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 the first

best. 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 the reclassification risk arising from 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, due to adverse selection, consumers' within-period event insurance unravels to the minimum allowable coverage of 60% actuarial value in equilibrium.⁴ 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 paths remains the same. However, with this lower level of risk aversion, dynamic contracts are preferred to the ACA-like market by consumers with all four income paths we consider, albeit the welfare losses from risk bearing are in all cases much smaller than in our base case (although not insubstantial). 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 increases the welfare optimal dynamic contracts deliver to the range of the ACA-like market. Third, we verify that our findings also hold using health state transitions computed for employees of a larger firm (so that data for some transition cells are less sparse). Fourth, we examine the degree to which optimal precautionary savings would improve welfare under spot contracting. While consumer welfare improves, the basic insights from our analysis are unchanged. Fifth, we examine the welfare optimal dynamic contracts deliver to consumers

⁴Risk adjustment, which we do not model here, could reduce this loss [see, for example, the discussion in Handel, Hendel, and Whinston (2015)].

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, so that all consumers have the same lifetime welfare as the healthiest 25-year old, would cost the government roughly \$2,000 per consumer. Alternatively, a revenue-neutral government tax/subsidy scheme that insures this pre-age 25 risk results in welfare similar to that in the ACA-like exchange.

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.⁵

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.⁶ We discuss both of these ideas, and their relation to our optimal contracts, in Section 7. 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 their 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 do not derive optimal contracts that take account of consumption-smoothing concerns, nor do they conduct a welfare analysis as we do.

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

⁵Abramitzky (2010) applies a similar model to understand the evolution and existence of the kibbutz as an institution.

⁶Pauly, Kunreuther, and Hirth (1995) refer to these policies as guaranteed renewable contracts, but (as they note) effectively treat them as guaranteed premium path contracts.

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.⁷

Compared to all of this previous work on long-term health insurance contracts, our work is unique in using data to derive optimal contracts with one-sided commitment, and in assessing the welfare impacts of these contracts. More generally, the paper provides an empirical assessment of the gains to optimal long-term contracting in an economically significant setting.

The rest of the paper proceeds as follows. Section 2 presents our model and its equilibrium implications. Section 3 describes 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 examines several extensions. Section 7 addresses 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, including a discussion of potential barriers to adoption of dynamic health insurance contracts.

2 Model of Dynamic Insurance

We consider a dynamic insurance problem T periods long, with periods indexed $t = 1, \dots, T$. 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. We take Λ to be a finite set. The consumer's initial health state is λ_1 , which occurs with probability $f_0(\lambda_1)$. 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 probability $f_t(m_t, \lambda_{t+1} | \lambda_t)$. We denote by $\bar{f}_t(\lambda_{t+1} | \lambda_t)$ the marginal conditional probability of λ_{t+1} . As future health is not relevant for the period T insurance problem, we need only specify the probability $f_T(m_T | \lambda_T)$.⁸

⁷Bundorf, 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.

⁸Incorporating 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. However, the optimal policy would then provide both health and income insurance, and could affect labor supply incentives.

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.⁹ We make the following assumptions concerning the stochastic health process:

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

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

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.

We assume that the insurance market is perfectly competitive, with risk-neutral firms who discount future cash flows using the discount factor $\delta \in (0, 1)$. A consumer's risk preferences are described by $u(\cdot)$, the consumer's Bernoulli utility function, while the consumer's long-run 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.¹⁰ Throughout we assume that consumers are unable to borrow to fund premium payments or other expenses.

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

2.1 Three Benchmarks

We will compare optimal dynamic contracts with one-sided commitment against three natural benchmarks.¹¹ 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

⁹Our assumption that all insurers have access to the same information assumes the ability of insurers to underwrite effectively potential new customers. If, instead, an individual's current insurer had better information than other firms, prospective insurers would face an adverse selection problem when attempting to attract lapsing consumers. For the consequences of this type of adverse selection, see, for example, DeGaridel-Thoron (2005).

¹⁰The 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 . As noted in footnote 8, in this case the optimal contract would insure both health and income risk.

¹¹We also compare the outcome of optimal dynamic contracts with one-sided commitment to an ACA-like exchange market institution.

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

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

As is well known, if consumers and insurance firms could both commit to a long-term contract prior to the realization of λ_1 , the competitive equilibrium would yield this outcome.

At the opposite extreme, long-term contracts may be impossible, leading to single-period “spot” insurance contracts. In a competitive market, 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, 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

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

Finally, in this dynamic setting both insurance and consumption smoothing over time are needed to achieve the first best. Since we will focus on settings in which income is increasing over time and borrowing is impossible, another natural benchmark is the outcome that would result if the consumer was fully insured within each period (eliminating all ex ante risk) but resources could not be transferred over time. This certain but time-varying consumption path results in the same welfare as the constant consumption level C_{NBNS}^* (“NBNS” = “No Borrowing/No Savings”) such that

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

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 assume that contracting begins in period 1 (in our empirical setting, at age 25) after λ_1 has been realized.¹² 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 through period t , $h^t =$

¹²We consider the effects of health risk before age 25 in Section 6.5.

$[\lambda_1, (m_1, \lambda_2), \dots, (m_t, \lambda_{t+1})]$.¹³ (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 (lapse would instead yield the initial insurer a continuation profit of zero). As a result, we can look for an optimal contract by imposing "lapse constraints" that require that after no history h^t is it possible to offer the consumer an alternative continuation contract that (i) itself prevents future lapse, (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 (which determines future expected medical expenses), and S_t is the absolute value of the loss that the insurer is allowed to sustain going forward (i.e., S_t is the *subsidy* for future insurance). This is a useful 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, given the consumer's current health state, is that it is not possible to increase the consumer's continuation utility while keeping the insurer's loss equal to S_t . So, the continuation of the contract must itself solve an optimal contracting problem for an insurer who can sustain the loss S_t starting in health state λ_t . Second, the constraint that the contract prevents lapse 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 :

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

¹³This 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.

In problem (4), the insurer offers $c_T(\cdot)$, which specifies a consumption level c_T for each realization of period T medical expenses m_T , subject to the constraint that the insurer's losses not exceed S_T . 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. 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 :

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

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, given the subsidy S_t that is available at the start of period t . The second constraint is the “lapse constraint,” which says that the consumer's continuation value for each realization of λ_{t+1} cannot fall below the value $V_{t+1}(\lambda_{t+1}, 0)$ 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 cannot 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:¹⁴

Proposition 1 *The equilibrium contract in a competitive market with one-sided commitment for a consumer of type $\theta = (y, u)$ and who cannot borrow is characterized by a collection of consumption guarantees $\{c_t^\theta(\lambda_t)\}$, where each $c_t^\theta(\lambda_t)$ is the consumption guarantee offered*

¹⁴Aside from the change of context from labor markets to health insurance, our model generalizes Harris and Holmstrom (1982) by allowing for more general stochastic medical expense and health state transition processes. We also offer a recursive proof of the result, show that the optimal contracts are self-selective (Section 2.3), and introduce switching costs (Sections 2.4 and 6.2).

to the consumer in the first period of a break-even (i.e., zero subsidy) contract starting in period t when the consumer is in health state λ_t . The consumer who agrees to a contract in period 1 when in health state λ_1 enjoys in each period t following health state history $(\lambda_1, \dots, \lambda_t)$ the certain consumption $\max_{\tau \leq t} c_\tau^\theta(\lambda_\tau)$. 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 $c_1^\theta(\lambda_1)$. This minimal guarantee is bumped up in later periods $t > 1$ to a new, higher guarantee $c_t^\theta(\lambda_t) > c_1^\theta(\lambda_1)$ to match the outside market at the first instance at which 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. That new guarantee is in turn bumped up in the periods that follow if, again, doing so is necessary to match the outside market. The equilibrium contract provides full within-period insurance for the consumer (i.e., consumption in each period is independent of m_t), and partial insurance against reclassification risk, as consumers who have experienced sufficiently bad health states leading up through a given period t [such that $c_\tau^\theta(\lambda_\tau) \leq c_1^\theta(\lambda_1)$ for all $\tau \leq t$] 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 $u(\cdot)$, as well as the health state transition process. Since the consumer's consumption level is always weakly rising over time, the consumer never wishes to save.

The optimal contract in Proposition 1 specifies directly the consumer's consumption levels and prevents lapsation. Note, however, that the same outcome can alternatively be achieved by means of a *guaranteed premium path contract* from which the consumer may lapse. Specifically, the consumer is given the option to renew, if she has not yet lapsed, at the guaranteed premium path $p^\theta \equiv (p_1^\theta, \dots, p_T^\theta)$ where $p_t^\theta = y_t - c_1^\theta(\lambda_1)$ for $t = 1, \dots, T$, provided that she has always renewed in the past.¹⁵ 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 $p_\tau^\theta = y_\tau - c_t^\theta(\lambda_t)$ for $\tau \geq t$, where $c_t^\theta(\lambda_t) > c_1^\theta(\lambda_1)$. 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.^{16,17}

¹⁵This form of contract is the counterpart to the "Annual Renewable Term" life insurance contracts studied in Hendel and Lizzeri (2003).

¹⁶We discuss in Section 8 the difference between this guaranteed guaranteed premium path contract and that of Pauly, Kunreuther, and Hirth (1995).

¹⁷The 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

2.3 Unobserved Types and Self-Selection

The analysis above assumed that a consumer's lifetime income path $y = (y_1, \dots, y_T)$ and risk preferences $u(\cdot)$, summarized in the consumer's "type" $\theta = (y, u)$, 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 collection of optimal contracts for all types derived above, presented as the guaranteed premium path contracts described above, consumers will self-select, choosing the optimal contract for their type.¹⁸

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

Our result is:

Proposition 2 *Suppose that, in each period $t = 1, \dots, T$, the menu of optimal guaranteed premium path contracts $\{p_t^\theta(\lambda_t)\}_{\theta \in \Theta, \lambda_t \in \Delta}$ is offered, where $p_t^\theta(\lambda_t) \equiv \{y_t - c_t^\theta(\lambda_t)\}_{t=1}^T$. Then in each period the menu is self-selective: that is, if a consumer of type θ agrees to a new contract she chooses that type's optimal contract $p_t^\theta(\lambda_t)$.*

Proof. In Appendix B. ■

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

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 $y_1 - c_1^\theta(\lambda_1)$, while the period 2 premium would equal $y_2 - \mathbb{E}[m_2|\lambda_2] - \max_{\tau \leq 2} c_\tau^\theta(\lambda_\tau) + S_2^*(m_1, \lambda_2)$. 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 $\max_{\tau \leq 2} c_\tau^\theta(\lambda_\tau)$. 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 into an insurer, perhaps resulting in better insurer performance as well as better matching of insurers and consumers when health care networks are bundled with insurance provision. However, such contracts require that courts can verify the consumer's health state λ , while guaranteed premium path contracts do not.

¹⁸Note 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.

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) expected medical costs conditional on an individual's health state, (ii) the transitions across health states as individuals age, (iii) preferences towards risk, and (iv) income paths.¹⁹

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 descriptive 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).

¹⁹One 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 expenses conditional on an individual's health state, rather than just its mean. See Handel, Hendel, and Whinston (2015).

3.1 Health States

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.²⁰ The output is an index that represents the health status of each individual in the population. Since the ACG is used by insurers in their underwriting processes, our empirics are based on similar information about risks that market participants (insurers) have.²¹

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.

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). For each age cohort, the first two columns show the 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 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 meaningful adverse selection if community rating is introduced (as under the Affordable Care Act), something that is shown in Handel, Hendel, and Whinston (2015).

²⁰We 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.

²¹This is one of the main advantages of our empirical strategy. Most of the literature on health insurance estimates the distribution of risks from observed insurance choices and realized total medical expenditures. Instead our measure of risk is based on diagnosis codes and professional diagnostics (the ACG index). The distribution of risk is data, as opposed to being inferred from choice and expense realizations.

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 average of the residual variation.

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, with bin 1 being the healthiest, and bin 7 being the sickest. 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 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. 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.²² 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.²³

²²Admittedly, by defining transitions over ACGs we may miss potential information on what condition led to the current ACG index that could entail different persistence.

²³To study the long run predictive power of these one-year ACG transitions we compared the proportion of individuals in each state at different ages in the sample, to the proportions predicted by the transition

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.

Health State Transitions: 30-35 Year Olds							
λ_t	1	2	3	λ_{t+1} 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.

Tables 3 and 4 present these transition matrices for ages 30-35 and 40-45 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 following 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 40-45 year olds, health states show similar persistence.

matrices for an individual who starts healthy at age 24. To assess the fit between predicted and actual proportions, we ran a regression of actual on predicted proportions. The coefficient on predicted proportions is 0.95 and the R-squared 0.77. Splitting the sample for the ages 25 to 44, and 45 to 64, the R-squared are 0.89 and 0.63. While the predictive power of the latter is lower, it involves predictions on average 30 years away, but is still quite a good fit.

Health State Transitions: 40-45 Year Olds							
λ_t	1	2	3	λ_{t+1} 4	5	6	7
$\lambda_t = 1$	0.57	0.27	0.08	0.01	0.01	0.04	0.02
$\lambda_t = 2$	0.21	0.37	0.14	0.14	0.05	0.05	0.04
$\lambda_t = 3$	0.08	0.16	0.29	0.18	0.10	0.13	0.07
$\lambda_t = 4$	0.09	0.14	0.19	0.17	0.20	0.12	0.09
$\lambda_t = 5$	0.05	0.10	0.24	0.23	0.10	0.12	0.16
$\lambda_t = 6$	0.04	0.14	0.09	0.20	0.18	0.09	0.27
$\lambda_t = 7$	0.00	0.03	0.03	0.17	0.20	0.20	0.37

Table 4: Health status transitions from one year to the next, for 40-45 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 the present value of expected health expenses for different periods (age 30, ages 31-35, 36-40, and 31-65), for an individual who at age 30 was in the respective health state shown in each row of the first column.

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 40. 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.^{24,25}

²⁴The fact that expected costs depend very little on the health state 10 years prior is consistent with actuarial mortality tables. There are two kinds of tables: “ultimate” tables are based on attained age only, while “select and ultimate” tables report the death rate not only by attained age, but by years since underwriting (namely, conditional on being in good health at that time). The tables converge as the years since underwriting increase; 10 years after underwriting the rates are quite similar.

²⁵In Appendix C we examine how the level of health state persistence affects welfare under optimal dynamic contracts relative to our benchmarks. One point to note is that complete persistence would eliminate the benefit of dynamic contracts as there would be no reclassification risk to insure once a consumer’s age 25 health state is realized.

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 the present value of expected health expenses for different periods (age 40, ages 41-45, 46-50, and 41-65), for an individual who at age 40 was in the respective health state shown in each row of the first column.

3.3 Risk Preferences

The third ingredient is a consumer’s risk aversion, i.e., the degree to which consumption smoothing over different states of the world is valued by consumers. 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:

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

where $c = y - p - o$ is consumption (which equals income less premium payments p and out-of-pocket medical expenses o) and γ is the risk aversion parameter.²⁶

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 our 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.

²⁶In 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 a discussion; a correlation test, in the spirit of Chiappori and Salanie (2000), does not reject this assumption.

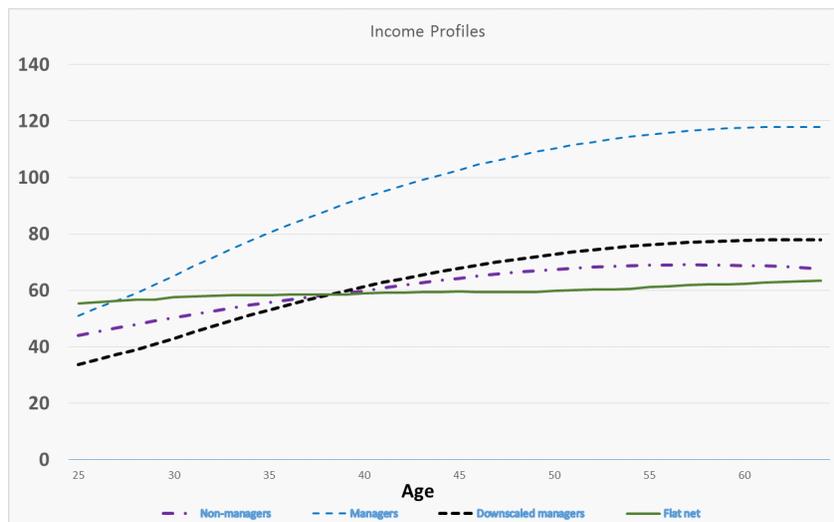


Figure 1: Income paths for managers, non-managers, proportionally “downscaled” managers, and flat net income.

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.²⁷ The latter eliminates intertemporal consumption smoothing motives (for an individual who would pay a certain premium in each period equal to her ex ante expected medical costs), as individuals with flat net income do not want to use the contract as a mechanism to borrow or save, unlike consumers with increasing or decreasing income profiles over time. 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 that grows more quickly than expected medical expenses, leading to steeper paths than the flat net income curve.

²⁷The manager and non-manager income paths we use here are estimated in Handel, Hendel, and Whinston (2015).

4 Computation of Optimal Dynamic Contracts

Given estimated values of transitions $f(\cdot)$ and preferences $u(\cdot)$, 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 $s \in \{1, \dots, 7\}$ is $c_\tau^y(\lambda_\tau = s)$, a consumption guarantee that breaks even for a contract starting in period τ with health state $\lambda_\tau = s$, given the future guarantees described in Proposition 1 (which are, 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 \bar{c} such that

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

where $P_\tau(z|\lambda_t = s, C_{t+1}^y, \bar{c})$ is the probability that, starting in health state $\lambda_t = s$ in period t , the health state transitions $(\lambda_{t+1}, \dots, \lambda_\tau)$ from period t to period τ are such that $\lambda_\tau = z$ and $c_{t'}^y(\lambda_{t'}) \leq \bar{c}$ for all $t' \in \{t+1, \dots, \tau\}$.²⁸ The first term in curly brackets in (7) is the expected profit in period t given the initial consumption guarantee \bar{c} , while the second term in curly brackets is the expected continuation profit and uses the fact that starting at the first instance at which the consumer's health state λ_τ is such that $c_\tau^y(\lambda_\tau) > \bar{c}$ (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.

²⁸These probabilities are computed using C_{t+1}^y and the transition matrices.

5 Results

Using the data and computational approach described above, we find the consumptions and premiums in optimal dynamic contracts. In what follows we first examine the structure of these contracts, and then turn to welfare analysis.

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 consumer 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. For example, 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 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. The least healthy, on the other hand, have very high costs (and, hence, high marginal utility of consumption) compared to their expected future costs (recall Table 5); for them, front-loading isn't worthwhile. 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 health state realizations.

Tables 8 and 9 present second-period premiums and consumptions, 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

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 front-loading, and consumption levels.

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.

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. Only when $\lambda_2 = 1$ does the lapsation constraint bind for this consumer, resulting in an increase in consumption to \$52,905 (and a corresponding reduction in her premium).

The lapsation constraint binds for more and more second-year states the sicker the consumer was at the start of the contract. For consumers initially in the sickest health 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.

Second-Year Equilibrium Consumptions: Flat Net Income								First-Year
λ_1	λ_2							Consumption
	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	51,145
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 consumptions in the equilibrium long-run contract for a flat net income path, as a function of the period 1 and period 2 health states.

Equilibrium Front-loading: Flat Net Income										
	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.

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 of the first ten years the mean front-loading (averaging across all possible histories, including the possible age 25 health states, weighted by the probability of the histories).²⁹ 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 \$2,536 in year 9, and declines to \$2,469 in year 10. Maximum possible front-loading increases slightly over time through year 9, and also declines in year 10.

²⁹Throughout our analysis, we assume that the consumer arrives at age 25 with the health distribution given by the transitions from the healthiest state for 25-30 year olds (i.e., we take the age-25 health distribution to be that generated by one transition from $\lambda = 1$ using the transition matrix in Table 3). The distribution of health states for 25-30 year olds implied by this assumption closely matches the distribution in Table 2, as discussed in footnote 22.

First-Year Equilibrium Contract Term: Downscaled Managers							
	1	2	3	λ_1 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 premiums, expected costs, the extent of front-loading, and consumption levels.

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.

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 early in life for long-term insurance quite costly, as marginal utility is high early compared to what is expected later in life. For example, a downscaled manager who is in the healthiest state at age 25 front-loads only \$547, compared to \$1,619 for a consumer with flat net income.

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 health 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 certainty equivalent represents the constant consumption for the forty years of life from age 25 to 65 that makes the consumer

Second-Year Equilibrium Premiums: Downscaled Managers								First-Year
λ_1	λ_2							Premium
	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 down-scaled manager income path, as a function of the period 1 and period 2 health states.

Second-Year Equilibrium Consumption: Downscaled Managers								First-Year
λ_1	λ_2							Consumption
	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 down-scaled manager income path, as a function of the period 1 and period 2 health states.

as well off as in a given market setup, ex ante.³⁰ Specifically, we compare the certainty equivalent of optimal dynamic contracts with one-sided commitment, denoted by CE_D , to the three benchmarks we have described previously (see Section 2.1 for formal definitions):

- (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 neither borrowing nor saving is possible, denoted C_{NBNS}^* .

We also compare optimal dynamic contracts with one-sided commitment to an alternative market design that has been 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 creating adverse selection for within-year coverage.

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

³⁰By “ex ante” we mean prior to learning the age 25 health state. (See footnote 29 for a discussion of how we determine that state.)

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

Column (2) shows welfare outcomes under spot contracts, with no protection for reclassification risk. As column (6) shows, the welfare loss from spot contracting, relative to the first best, is highest for relatively steep income profiles (40.9% for downscaled managers). With steep income profiles, the first best both eliminates reclassification risk, and smooths consumption over time. However, the welfare loss from spot contracts is substantial (13.8%) even in the case of flat net income, where no intertemporal consumption smoothing is occurring in the first best.³¹

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, this may be a more relevant benchmark of how well optimal dynamic contracts with one-sided commitment do at eliminating reclassification risk, since saving and borrowing on their own can greatly improve utility for steep net income paths. Column (8) therefore reports $\frac{CE_D - CE_{SPOT}}{C_{NBNS}^* - CE_{SPOT}}$, which shows the proportion of the gap between spot contracting and this No-borrowing/No-saving benchmark that is closed by optimal dynamic contracts. 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. In sum, dynamic contracts are very effective at eliminating reclassification risk when net income paths are fairly flat, but with rising income paths they are only partially effective.

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 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

³¹In Section 6.4 we extend this analysis to allow precautionary savings in the spot contracting 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}	$\frac{C^* - CE_{SPOT}}{C^*}$	$\frac{CE_D - CE_{SPOT}}{C^* - CE_{SPOT}}$	$\frac{CE_D - CE_{SPOT}}{C_{NBNS}^* - CE_{SPOT}}$
Flat net	53.67	46.27	52.77	53.53	51.30	13.8%	87.7%	89.5%
Non-mngr	53.67	40.73	44.10	47.39	46.25	24.1%	26.0%	50.6%
Manager	84.00	50.32	51.77	56.08	55.09	40.1%	4.3%	25.2%
Downs Mngr	53.67	31.74	34.10	37.93	36.84	40.9%	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.

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 consumers' risk tolerance for the larger stakes associated with reclassification risk.³² The scale of losses from reclassification risk is larger in our environment than in the setting in which risk preferences were estimated.

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

³²See Rabin (2000) for a discussion of issues with CARA and scaling of gambles.

Income	Certainty Equivalent (\$1,000s)					Gains from Long-term contracts		
	C^*	CE_{SPOT}	CE_D	C_{NBNS}^*	CE_{ACA}	$\frac{CE_D - CE_{SPOT}}{C^* - CE_{SPOT}}$	$\frac{C_{NBNS}^* - CE_{SPOT}}{C_{NBNS}^*}$	$\frac{CE_D - CE_{SPOT}}{C_{NBNS}^* - CE_{SPOT}}$
Flat-net	53.67	52.47	53.62	53.67	52.85	95.8%	2.2%	95.8%
Non-mngr	53.67	50.91	51.70	51.84	51.35	28.6%	1.8%	85.0%
Manager	84.00	68.73	69.06	69.39	69.00	2.2%	1.0%	50.0%
Downs mngr	53.67	46.41	46.94	47.20	46.80	7.3%	1.7%	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.

associated with this risky prospect. Instead, the lower risk aversion 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 between 1.0% and 2.2%, depending on the income path. For the higher 0.0004 risk aversion, the loss was between 13% and 41%. Still, a loss of 1-2% of lifetime certainty equivalent is not insubstantial.

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 (although, as noted above, the losses due to risk bearing are much smaller than before in both cases).

6.2 Switching Costs

Recent evidence from health insurance markets [Handel (2013), Ho et al. (2016)] points to substantial inertia in insurance choice. Switching costs have the potential to improve how dynamic contracts perform. Our basic model assumes that consumers lapse whenever they get a better offer. As discussed in Section 2.4, switching costs relax the lapsation constraints, which can enhance commitment and the welfare achievable with optimal dynamic contracts.

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 and risk aversion of 0.0004 for different levels of switching costs and four income profiles.

Table 16 shows the welfare achieved by dynamic contracts for different levels of switching costs and our four income profiles. As expected, 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 path and switching costs of \$5,000 and \$1,000. It is interesting to note that for all second-year states without a binding lapsation constraint consumption is higher with a higher switching cost, while consumption is lower for second-year states with a binding lapsation constraint.

First and Second-Year Consumptions: Switching Cost = 5,000								First Year
λ_1	1	2	3	λ_2 4	5	6	7	Consump
1	53,612	53,612	53,612	53,612	53,612	53,612	53,612	53,612
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 Consumptions: Switching Cost = 1,000								
1	2	3	4	5	6	7	8	
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	37,759	37,759

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

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. While our transition matrices, such as Tables 3 and 4, largely satisfy the stochastic dominance property (Assumption A2), there are a few violations of this property.³³ The violations probably reflect the sparsity of individuals in some cells. 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.³⁴

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

³³The assumption is sufficient for our characterization result (Proposition 1), but not necessary.

³⁴We 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 (1,000s)			Gains from Long-term contracts	
	C^*	CE_{SPOT}	CE_D	$\frac{C^* - CE_{SPOT}}{CE_{SPOT}}$	$\frac{CE_D - CE_{SPOT}}{C^* - CE_{SPOT}}$
Income					
Flat net	51.22	44.60	49.98	14.8%	81.3%
Non-mngr	51.22	38.14	41.01	34.3%	21.9%
Manager	81.55	47.37	48.72	72.2%	4.0%
Downs Mng	51.22	28.87	31.01	77.4%	9.6%

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

can rescale the numbers in Table 18 so that C^* equals that in the analysis above; after this adjustment, all of the welfare numbers with the alternative transitions are within 3% of our base numbers, reported in Section 5.3.³⁵

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 also would not want to engage in savings in the first best. However, with spot contracting consumers may want to engage in precautionary savings to lower the costs of reclassification risk. Individuals can save in good states to weather periods of bad health. To study the impact of precautionary savings we solve a finite-horizon savings problem, with the same underlying fundamentals as in our main analysis, namely, the same income paths, risk preferences, and transition matrices. We find optimal savings starting at age 25 given an income path and the actuarially fair health insurance premiums associated with the different health states.³⁶ 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”).

³⁵Note 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.

³⁶For each income profile, we solve a finite-horizon dynamic programming problem, from ages 25 to 65. Starting at age 64, for a grid of saving values entering that period, the individual finds the optimal saving level going into the last period that maximizes the sum of current utility from consumption and the discounted value of the expected utility in the last period, where the expectation is taken for each state given the transition matrices. Once we obtain the value function at age 64 for each possible health state and incoming saving level, we proceed backwards all the way to age 25, where we obtain the discounted expected utility starting in each possible health state. The ex-ante certainty equivalent is the certain consumption level that makes the consumer indifferent to the expected utility of entering the dynamic problem before observing the health realization at age 25.

	Certainty Equivalent (1,000s)			% Gains		
	(1)	(2)	(3)	(4)	(5)	(6)
Income	CE_D	CE_{SPOTws}	CE_{SPOT}	$\frac{CE_D - CE_{SPOT}}{CE_{SPOT}}$	$\frac{CE_{SPOTws} - CE_{SPOT}}{CE_D - CE_{SPOT}}$	$\frac{CE_D - CE_{SPOTws}}{CE_{NSNB} - CE_{SPOTws}}$
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 equating consumption across all states in which the lapsation does not bind. As well, 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 precautionary savings outcome that is closed by optimal dynamic contracts. As in our earlier analysis, dynamic contracts close a significant fraction of this gap with flat net income, but 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.

Table 20 presents welfare when contracting starts in various health states at age 25, for a non-manager income path. Starting in a good initial state has a limited impact on first-best welfare because of the relatively temporary nature of a poor initial health state. However, it has a dramatic effect on the welfare achieved under optimal dynamic contracts with one-sided commitment because starting in a poor health state makes front-loading much less attractive. As a result, there is a dramatic effect of the age-25 health state of a consumer on

Welfare and Initial Health (\$1,000s): Non-managers						
Initial state						
	C^*	CE_D	CE_{SPOT}	CE_{SPOTwS}	Subsidy	Break-even subsidies
Ex ante	53.67	44.10	40.73	42.19		46.70
λ_1 :						
1	53.76	47.20	41.79	43.70	–	–1.74
2	53.56	46.05	40.82	42.71	4.20	1.79
3	53.35	44.55	39.36	41.21	9.10	6.47
4	53.29	44.34	39.23	41.05	10.27	7.57
5	53.08	42.10	37.37	39.25	15.62	12.88
6	52.86	39.47	37.26	37.87	21.07	18.27
7	52.50	33.16	33.04	33.10	30.43	27.69
$E[\cdot]$					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.

her lifetime welfare under this regime. The situation with spot contracts (with or without precautionary savings) is intermediate.

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 that arises from uncertain age-25 health, perhaps through some form of government insurance. The next-to-last column of Table 20 shows the subsidy that is required in each age-25 health state to give a consumer facing the dynamic contracting environment the same welfare as if they were in the healthiest age-25 state.³⁷ 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. The last column instead reports the expected revenue-neutral age-25 taxes and subsidies that would insure this risk, and (at the top of the column) the resulting certainty equivalent of \$46,704. With this balanced-budget government scheme, the welfare from dynamic contracts slightly exceeds that in the (non-risk-adjusted) ACA-like market for these consumers (reported in Table 14).

³⁷For each state we find the subsidy that, when added to the first year of the consumer’s income path, leads to a certainty equivalent of \$47,200 with an optimal dynamic contract.

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, PKH aim to design a policy that provides full insurance in each period and guarantees that the consumer can renew in the future at the same premium as would be offered to the healthiest consumer type at that age. The idea is that a consumer with such a policy never wishes to lapse and faces no uncertainty in their consumption (i.e., 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):³⁸

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

To understand this formula, consider a case in which $T = 2$. In period 2 (the last period), the consumer pays a premium equal to $p_2 = \mathbb{E}[m_2 | \lambda_2 = 1]$, 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 and risk preferences. The PKH contracts go to the extreme of completely preventing reclassification risk, resulting in a fully deterministic

³⁸We provide the derivation in Appendix D.

consumption path but excessively low initial consumption.³⁹ 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: the healthiest types would actually be drawn away by a firm offering the contracts described in Proposition 1.⁴⁰

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 her premium.⁴¹ In each period t the consumer would pay a premium of

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

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

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

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 the period t premium (9), we see that the net premium payment in each period t for a consumer who begins with $\lambda_1 = 1$ is exactly the PKH premium (8).

³⁹PKH'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)].

⁴⁰This lapsation problem in the PKH contract arises only if $T \geq 3$.

⁴¹Cochrane (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.

7.1 Empirical Comparison of PKH and Optimal Dynamic Contracts

Using formula (8), we calculate that the initial PKH premium paid by a healthy 25-year old is about 27% higher than the initial premium paid by a healthy 25-year old individual with flat net income in the optimal dynamic contract.^{42,43} For a consumer with a flat net income path, the excessively low initial 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 relative 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 16.6% compared to an optimal dynamic contract, while for a non-manager $CE_{PKH} = 42,758$ which represents a 16.0% welfare loss.

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 characterized optimal contracts, and quantified their positive and normative implications relative to alternative market designs using detailed individual-level data on health risks, health state transitions, insurance choices, and income profiles for the employees of a large firm.

Optimal dynamic contracts exhibit front-loading at early ages, thereby enabling consumption guarantees that reduce reclassification risk. These guarantees can also be provided by offering an equivalent guaranteed premium path whose year-to-year increases match the consumer's income growth. Moreover, when framed as guaranteed premium path contracts, these optimal contracts – which differ according to a consumer's income path and risk preferences – are self-selective, in the sense that consumers with differing types will choose the contract that is designed for them.

Empirically, we find that optimal dynamic contracts for our population of consumers exhibit substantial front loading: at age 25 a healthy consumer with a flat net income path, for example, pays a premium of \$2,750 compared to expected medical costs of only \$1,131.

⁴²In 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.

⁴³At age 25 the value of the second term in equation (8), representing the premium pre-payment that is required in the PKH contract, is \$4,604. This amount divided by δ ($= 0.975$) is also the end-of-period amount that the consumer would need to pay out in the event that she remained healthy (with $\lambda_{26} = 1$) to achieve the first best in the reclassification-risk insurance scheme proposed by Cochrane (1995).

However, a healthy “downscaled manager,” whose 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: Optimal dynamic contracts with one-sided commitment almost achieve the first-best for consumers with flat net income paths, but offer less improvement over 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 considered 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 contracting institutions, with low risk aversion dynamic contracts offer higher welfare than an ACA-like market for all income paths. (Of course, the differences in welfare are smaller when risk aversion is lower.) 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. Alternatively, a revenue-neutral government tax/subsidy scheme can insure this pre-age-25 risk, resulting in welfare similar to that in an ACA-like exchange.

In sum, dynamic contracts offer a way to reduce reclassification risk without requiring community rating as in the ACA. However, with rising incomes we find in our base specification that the ACA approach (even without effective risk adjustment) offers greater lifetime welfare. This conclusion can be reversed, though, if consumers’ risk aversion is lower (in which case risk-bearing has much smaller, though not insubstantial, welfare consequences), consumers face significant switching costs, or the government insures pre-age-25 (i.e., pre-contracting) risk.

A natural question, then, is whether there are other factors that might in practice limit the use of dynamic contracts, but that we have abstracted from. We conclude with a discussion of this issue.

Our model has abstracted from a number of factors that could potentially reduce the benefits of dynamic contracts. One concern is that firms may have difficulty forecasting future medical cost levels, an issue that does not arise in markets such as life insurance

in which long-term contracts are prevalent. Moreover, this risk is not fully diversifiable.⁴⁴ Nonetheless, this need to forecast can be alleviated (in part) by indexing future guaranteed premiums to medical cost inflation indices, which also would allow for a sharing of the risk between firms and consumers.

Another potential problem is that consumer lock-in might lead to quality degradation by insurers. However, it would seem that insurers' quality incentives would actually be enhanced in some dimensions, as they would have increased incentives to promote long-term health. Lock-in could also reduce a consumer's ability to re-match with firms if firm-specific preferences change (Atal (2015)).⁴⁵

One might also be concerned that front-loaded contracts would be unappealing to "young invincibles," a group that has been difficult to attract into the ACA market. However, for these consumers (i.e., 25 year old consumers with $\lambda_{25} = 1$) the optimal contract's \$2,750 premium for the case of flat net income is less than the average medical cost for 25-30 year olds (equal to \$3,112), which would be the premium in an age-based community rating environment with full insurance. For healthy 25-year-old downscaled managers the optimal contract premium is even less: \$1,678.

Perhaps the most serious current limitation on the benefits of dynamic contracts is that consumers have a tax-based incentive to use employer-based insurance when it is available to them, which can limit the duration of consumers' involvement in the individual market. For example, in the pre-ACA world, while some consumers purchased individual insurance over long periods of time, many others used it as a short-term solution between employment spells, leading median duration in the individual market in one study to be less than 2 years [Marquis et al (2006); see also Herring, Song, and Pauly (2008)]. Short durations can greatly reduce the benefits of a long-term contract.⁴⁶ In addition, those older consumers newly arriving to the individual market with pre-existing conditions (perhaps because of a job loss) would still face reclassification risk, much as in our discussion of unhealthy 25-year-old consumers in Section 6.5, perhaps necessitating some sort of government insurance (such as high-risk pool subsidies).

Another way to address this question is to look at the extent to which dynamic contracts are used in practice. In the U.S., before the ACA mandated annual policies, few explicitly

⁴⁴The need to forecast could also introduce "winner's curse" type concerns, as firms who attract a lot of business would tend to be those whose forecasts of future medical cost inflation are unreasonably low.

⁴⁵This problem would be greatly reduced if health insurance products were purely financial.

⁴⁶Introducing an exogenous probability of break up into our model is equivalent to lowering the discount rate, provided that separation payments upon break up cannot be made. With a discount factor of 0.5 (which generates an expected duration of 2), optimal dynamic contracts close only 36.8% of the gap between spot contracts and the no-borrowing/no-savings benchmark for a consumer with flat net income.

dynamic contracts seem to have been offered.⁴⁷ Nonetheless, it does appear that there were important dynamic aspects to the contracts offered to consumers. While good evidence on the workings of the pre-ACA individual market is limited, it appears that in most states insurers faced guaranteed renewability regulations that prevented them from re-pricing a policy to continuing customers on an individual basis [Patel and Pauly (2002)].⁴⁸ There is also evidence that these constraints, where present, did in fact significantly limit reclassification risk consumers faced once enrolled in a policy [Marquis and Buntin (2006), Herring and Pauly (2006), Herring, Song, and Pauly (2008)].^{49,50} At the very least, this suggests that insurers can deal with considering dynamic concerns in their pricing of policies. However, why insurers did not offer optimal dynamic contracts to improve upon the regulated guaranteed renewable policies is an open question; whether these regulated contracts were close to optimal, whether instead insurance regulations made it difficult to do so, or whether other impediments existed (such as the short average duration of consumers in the individual market) is not clear.

Of course, the high number of uninsured consumers in the pre-ACA world is well-documented. It is not clear that optimal dynamic contracts will significantly help with this problem, especially given that some form of dynamic contracting seems to have been part of the pre-ACA environment. In part, this high rate of uninsurance may have been due to problems of adverse selection [Hendren (2013)] and the availability of free care through various means [Mahoney (2015)]. As well, the poor may have a low demand for high-cost medical expenditures given their limited wealth, and hence also a low demand for insurance that covers such care. As a result, significant expansion of comprehensive insurance, whether via dynamic contracts or annual contracts, may require large government subsidies. What our analysis shows is that significant insurance of reclassification risk may be effectively provided via long-term contracts that are risk-rated, rather than through short-term community-rated policies, provided that consumers enroll when young and healthy.

⁴⁷One attempt to do so is described in <http://www.nytimes.com/2008/12/03/business/03insure.html> (accessed February 23, 2017). The possible adoption of the ACA may have reduced demand for this policy.

⁴⁸Insurers were generally free to set initial premiums based on health status (Kaiser Family Foundation (2012)), and could also raise a policy's premiums on a non-individual basis, although potentially subject to state-level regulatory constraints. The pool of customers in a policy generally worsened over time due to attrition of the healthiest consumers (much as in our theory) and insurers would raise rates over time to the continuing pool of customers (so-called "durational rating"). This ability to raise rates did allow insurers to avoid medical cost inflation risk, as a policy's premiums could be increased in response to cost increases.

⁴⁹However, consumers with pre-existing conditions needing to insure for the first time in the individual market (perhaps because of a job loss) faced risk-rating in most states.

⁵⁰Similar regulations are present in Germany and Chile, as noted in Browne and Hoffman (2013) and Atal (2016), respectively. Germany requires premiums to be constant over time, although they are indexed, much as we discussed above. In Chile firms face restrictions on how they initially risk rate and also a band limiting the range of policy-level price increases, although average policy-level price increases are unregulated.

References

- [1] Abramitzky, R.(2008) “The Limits of Equality: Evidence from the Israeli Kibbutz,” *Quarterly Journal of Economics* 123: 1111-59.
- [2] Akerlof, G. (1970) “The Market for ‘Lemons’: Quality Uncertainty and the Market Mechanism,” *Quarterly Journal of Economics* 84: 488-500.
- [3] Atal, J. P. (2016), “Lock-In in Dynamic Health Insurance Contracts: Evidence from Chile,” working paper.
- [4] 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] Chiappori, P. and B. Salanie (2000), “Testing for Asymmetric Information in Insurance Markets,” *Journal of Political Economy* 108: 56-78.
- [10] Cochrane, J. (1995), “Time-Consistent Health Insurance,” *Journal of Political Economy* 103: 445-73.
- [11] DeGaridel-Thoron, T. (2005), “Welfare-Improving Asymmetric Information in Dynamic Insurance Markets,” *Journal of Political Economy* 113: 121-50.
- [12] Finkelstein, A, K. McGarry and A. Sufi. (2005) “Dynamic Inefficiencies in Insurance Markets: Evidence from Long-Term Care Insurance,” *The American Economic Review* 95: 224-8.

- [13] Fudenberg, D., B. Holmstrom, and P. Milgrom (1990), “Short-term Contracts and Long-term Agency Relationships,” *Journal of Economic Theory* 51: 1-31.
- [14] Hendren, N. (2013), “Private Information and Insurance Rejections,” *Econometrica* 81: 1713-62.
- [15] Herring, B. and M.V. Pauly (2006), “Incentive-Compatible Guaranteed Renewable Health Insurance Premiums,” *Journal of Health Economics* 25: 395-417.
- [16] Ho, K., J. Hogan, and F. Scott Morton (2016), “The Impact of Consumer Inattention on Insurer Pricing in the Medicare Part D Program,” NBER working paper no. 21028.
- [17] 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.
- [18] Handel, B. (2013), “Adverse Selection and Switching Costs in Health Insurance Markets: When Nudging Hurts,” *American Economic Review* 103: 2643-82.
- [19] Handel, B., I. Hendel and M. Whinston (2015), “Equilibria in Health Exchanges: Adverse Selection vs. Reclassification Risk,” *Econometrica* 83: 1261-1313.
- [20] Harris, M. and B. Holmstrom (1982), “A Theory of Wage Dynamics,” *Review of Economic Studies*, 40: 315-33.
- [21] Hendel, I. (2016), “Dynamic Selection and Reclassification Risk: Theory and Empirics,” working paper.
- [22] Hendel, I. and A. Lizzeri (2003), “The Role of Commitment in Dynamic Contracts: Evidence from Life Insurance,” *Quarterly Journal of Economics* 118: 299-327.
- [23] Herring, B. and M. Pauly (2006), “Incentive-Compatible Guaranteed Renewable Health Insurance Premiums,” *Journal of Health Economics* 25: 395-417.
- [24] Herring, B., X. Song, and M. Pauly (2008), “Changes in Coverage in the Individual and Group Health Insurance Markets and the Effect of Health Status,” U.S. Department of Health and Human Services Report, Office of the Assistant Secretary for Planning and Evaluation.
- [25] 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.

- [26] Koch, T. (2011), “One Pool to Insure Them All? Age, Risk, and the Price(s) of Medical Insurance,” working paper.
- [27] Marquis, M.S. and M.B. Buntin (2006), “How much risk pooling is there in the individual insurance market?,” *Health Services Research* 41: 1782-1800.
- [28] Marquis, M.S., M.B. Buntin, J.J. Escarce, K. Kapur, T.A. Louis, and J.M. Yegian (2006), “Consumer Decision Making in the Individual Health Insurance Market,” *Health Affairs* 25: w226-34.
- [29] Mahoney, N. (2015), “Bankruptcy as Implicit Health Insurance,” *American Economic Review* 105: 710-46.
- [30] Pauly, M., H. Kunreuther and R. Hirth (1995), “Guaranteed Renewability in Insurance,” *Journal of Risk and Uncertainty* 10: 143-56.
- [31] Pauly, M., A. Percy, and B. Herring (1999), “Individual vs. Job-Based Health Insurance: Weighing the Pros and Cons,” *Health Affairs* 18: 28-44.
- [32] Rabin, M. (2000), “Risk Aversion and Expected-Utility Theory: A Calibration Theorem,” *Econometrica* 68: 1281-92.
- [33] 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 $\sigma \geq 0$ that represents the consumer’s consumption loss should she switch. 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 for a consumer of type $\theta = (y, u)$ who cannot borrow and has a switching cost $\sigma \geq 0$ is characterized by a collection of consumption guarantees $\{c_t^{\theta\sigma}(\lambda_t)\}$, where each $c_t^{\theta\sigma}(\lambda_t)$ is the consumption guarantee offered to the consumer in the first period of a break-even (i.e., zero subsidy) contract starting in period t when the consumer is in health state λ_t . The consumer who agrees to a contract in period 1 when in health state λ_1 enjoys in each period t following*

health state history $(\lambda_1, \dots, \lambda_t)$ the certain consumption $\max_{\tau \leq t} c_\tau^{\theta\sigma}(\lambda_\tau)$. Consumers have no incentive to save after agreeing to the 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 (i.e., the available *subsidy*), 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 $\bar{V}_T(\lambda_T, S_T, \sigma_T)$:

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

The FOC for this problem is

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

So we have

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

and

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

Observe, that we can rewrite problem (11) as

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

and in turn as

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

which shows that $\bar{V}_T(\lambda_T, S_T, \sigma_T) = \bar{V}_T(\lambda_T, S_T - \sigma_T, 0)$. 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 subsidy of $S_T - \sigma_T$.

It will be convenient to define the value function $V_T(\lambda_T, S_T - \sigma_T) \equiv \bar{V}_T(\lambda_T, S_T - \sigma_T, 0)$, 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 $\bar{V}_t(\lambda_t, S_t, \sigma_t)$:

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

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 $(\lambda_t, S_t, \sigma_t)$ 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

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

Thus, we have:

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

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

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

Proof. From Lemma 1, the inequality $\bar{V}_{t+1}(\lambda_{t+1}, S_{t+1}, 0) \geq \bar{V}_{t+1}(\lambda_{t+1}, 0, \sigma)$ is equivalent to the inequality $\bar{V}_{t+1}(\lambda_{t+1}, S_{t+1}, 0) \geq \bar{V}_{t+1}(\lambda_{t+1}, -\sigma, 0)$, 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 $V_t(\lambda_t, S_t - \sigma_t)$:

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

Remark 3 *We write the second constraint with $\delta f_t(m_t, \lambda_{t+1} | \lambda_t)$ (which is strictly positive) to simplify statement of the FOC (17) below. We also write the budget constraint as an equality, as it will clearly bind at the solution.*

9.3 Equilibrium Contract Characterization

Since without loss of generality the equilibrium contract prevents lapsation, it 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:

$$u'(c_t(m_t, \lambda_{t+1})) = \gamma_t \text{ for all } (m_t, \lambda_{t+1}) \tag{16}$$

$$\frac{\partial V_{t+1}(\lambda_{t+1}, S_{t+1}(m_t, \lambda_{t+1}))}{\partial S_{t+1}} = \gamma_t - \mu_t(m_t, \lambda_{t+1}) \text{ for all } (m_t, \lambda_{t+1}) \tag{17}$$

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:

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

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

Lemma 3 $S_{t+1}(\cdot)$ 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 $\bar{f}_t(\lambda_{t+1}|\lambda_t)$ be the marginal conditional distribution of λ_{t+1} given λ_t , we can rewrite the period t problem (for $\sigma_t = 0$) as

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

Its FOC are (19) and

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

Lemma 4 $S_{t+1}(\cdot)$ 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). ■

The next three results show how the consumption and the subsidy varies with the state (λ_t, S_t) (only the last of the three is necessary for the proof of Proposition 3).

Lemma 5 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 6 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 5 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 7 c_t is strictly increasing in λ_t , and γ_t is strictly decreasing in λ_t .

Proof. Suppose that $\lambda'_t > \lambda_t$ but $\gamma'_t \leq \gamma_t$. Denote the solutions for (λ_t, S_t) and (λ'_t, S_t) by $[c_t, S_{t+1}(\cdot)]$ and $[c'_t, S'_{t+1}(\cdot)]$ respectively. Then, by (21) and Lemma 2, 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 4, then imply that

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

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 argue 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).⁵¹

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 $p = (p_\tau, \dots, p_T)$ and $\hat{p} = (\hat{p}_\tau, \dots, \hat{p}_T)$ be guaranteed premium paths (that start in period τ) such that $\hat{p} \geq p$. Suppose that (i) the insurer earns a non-negative expected continuation payoff when guaranteed premium path p is chosen in period τ by a type θ consumer in health state λ_τ (given the consumer's optimal lapsation behavior), and (ii) under premium path p , in every period $t > \tau$ and health state λ_t in which a type θ consumer optimally does not lapse, the insurer's expected continuation payoff is non-positive. Then the insurer's expected continuation payoff is non-negative when premium path \hat{p} is chosen by a type θ consumer in health state λ_τ .*

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

Note, first, that $U(t, \lambda_t) \geq \hat{U}(t, \lambda_t)$: starting in period t , the consumer who faces p could adopt the same lapsation behavior as when facing \hat{p} and receive a weakly higher continuation payoff since under p she would be facing lower prices. Her optimal lapsation behavior yields a still higher payoff. Next, the fact that $U(t, \lambda_t) \geq \hat{U}(t, \lambda_t)$ implies that $\Lambda(t) \subseteq \hat{\Lambda}(t)$: in any health state in which the consumer lapses in period t when facing p , she also lapses when facing \hat{p} . Finally, consider the expected payoff of the insurer starting at state (τ, λ_τ) under p . This is the probability weighted average of the payoffs achieved along the various possible sequences of health states $(\lambda_\tau, \dots, \lambda_T)$. For each sequence the insurer earns premiums and

⁵¹Equivalently, 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 .

incurs costs until the consumer lapses. Since $\Lambda(t) \subseteq \widehat{\Lambda}(t)$, each such sequence hits lapsation weakly earlier under \widehat{p} than under p . Since, under path p , $\Pi(t, \lambda_t) \leq 0$ if $t > \tau$ and $\lambda_t \in \Lambda_{NL}(t)$, the earlier termination behavior under $\widehat{\Lambda}$ (but earning the same premiums p prior to lapsation) would weakly raise the expected payoff earned by the insurer for the sequence by changing a non-positive expected continuation payoff into a continuation payoff of zero. Moreover, the fact that the premiums earned until lapsation are higher under \widehat{p} than under p , while the expected costs are the same, means that a change from premium path p to path \widehat{p} , holding lapsation behavior fixed at $\widehat{\Lambda}$, would further raise the insurer's expected payoff earned from this health state sequence. As a result, $\Pi(\tau, \lambda_\tau) \leq \widehat{\Pi}(\tau, \lambda_\tau)$. ■

We next establish the following Lemma:

Lemma 10 *Suppose that in each period $t \geq \tau$ the menu of contracts offered to a consumer who is in health state λ_t and wishes to sign a new contract is the set of optimal contracts for that consumer, $\{p_t^\theta(\lambda_t)\}_{\theta \in \Theta, \lambda_t \in \Delta}$, and that moreover, in each period $t > \tau$ this menu is self-selective. Then an insurer earns a non-negative continuation expected NPV if in period τ type θ' consumer in health state λ_τ chooses the guaranteed premium path $p_\tau^{\theta'}(\lambda_\tau)$ that is intended for a type θ consumer in health state λ_τ .*

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

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

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

Now suppose that the Induction Hypothesis holds for periods t, \dots, T , and consider period $t - 1$ ($\geq \tau + 1$) after some previous history of health states and a sequence of decisions in which the insured has not yet lapsed. Suppose first that, under $p_\tau^{\theta'}(\lambda_\tau)$, period $t - 1$ lapsation behavior is different for type θ' than for type θ in a health state λ_{t-1} in which type θ would not lapse. Then since the lapsation constraint in problem (20) requires that $S_{t-1}(\lambda_{t-1}) \geq 0$ in the solution to the type θ problem in period $t - 1$, lapsation by type θ' in state λ_{t-1} would remove a continuation that had a weakly negative continuation expected payoff for the insurer when facing type θ and replace it with a zero payoff when facing type θ' .

Suppose, instead, that state λ_{t-1} is one in which type θ would lapse in period $t - 1$, choosing a contract with premium path \widehat{p} , while type θ' does not lapse. We will show that this changes what would have been a zero payoff continuation for the insurer into a continuation with a non-negative expected payoff when facing type θ' . By the self-selection assumption, we know that the contract \widehat{p} type θ chooses is $p_{t-1}^\theta(\lambda_{t-1})$, the optimal guaranteed premium path contract for that consumer, so the insurer offering that contract breaks even. Note now that since that contract induces the type θ consumer to lapse there is a $\Delta > 0$ such that $\widehat{p}_k = p_k^\theta - \Delta$ for all periods $k \geq t - 1$ (this is true because the two guaranteed premium paths differ only in offering different initial premiums and then the premium change each period equals the change in the type θ 's income). Hence, by Lemma 9, if the type θ consumer were instead not to lapse from path $p_\tau^\theta(\lambda_\tau) \equiv (p_\tau^\theta, \dots, p_T^\theta)$ in this state, the insurer's expected continuation payoff would be non-negative. But the Induction Hypothesis then implies that it is also non-negative when the type θ' consumer does not lapse in this state: the insurer's payoffs in period $t - 1$ from the two types are the same as both the premium paid and the expected medical costs are the same for the two types. The transitions to the period t state λ_t are also the same. But, by the Induction Hypothesis, the insurer's expected continuation payoff under contract $p_\tau^\theta(\lambda_\tau)$ is weakly higher starting in period t when facing the type θ' consumer than when facing a type θ consumer. So the Induction Hypothesis holds in period $t - 1$, and hence – applying induction – in period 2.

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

Apply induction ■

We now prove Proposition 2:

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

Induction Hypothesis: In each period $t > \tau$ the menu is self-selective: that is, if a consumer of type θ agrees to a new contract she chooses that type's optimal contract $p_t^\theta(\lambda_t)$.

The hypothesis is clearly true for $\tau = T$, as given any previous history the menu $\{p_T^\theta(\lambda_T)\}$ is a singleton with $p_T = \mathbb{E}[m_T|\lambda_T]$, and hence necessarily self-selective. Now suppose it is true for some τ ; we argue that it is then also true for $\tau - 1$. From Lemma 10 we know that if a type θ consumer chooses in period $\tau - 1$ when in health state $\lambda_{\tau-1}$ the contract intended for her then the insurer breaks even, but if she chooses instead the contract intended for type θ' the insurer earns a non-negative profit. But the policy intended for the type θ consumer maximizes the type θ consumer's discounted expected utility subject to the constraint that the insurer at least break-even. The policy intended for type θ' was therefore feasible for type θ , which implies that it cannot be preferred by type θ .

Apply induction. ■

Remark 4 *The proof goes through unchanged if we also allow consumers to have unobservable risk aversion coefficients. It does not go through, however, if consumers have unobservable switching costs. This makes sense: As discussed in Sections 2.4 and 6.2, high switching costs can, for example, allow the consumer to borrow, but a low-switching cost consumer would then find it worthwhile to misrepresent his type, borrow and then lapse.*

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 of past health. Note also that, with this transition process, ex ante (prior to the realization of λ_1) 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 precautionary savings possible (CE_{SPOTws}).⁵² (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

⁵²Precautionary savings was discussed in Section 6.4.

ρ	C^*	C_{SPOT}	CE_D	C_{SPOTwS}	$\frac{CE_D}{CE_{SPOT}}$	$\frac{CE_D}{CE_{SPOTwS}}$
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.

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 by optimal dynamic contracts with one-sided commitment 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 monotonically in ρ , while it is inverse U-shaped when there is precautionary savings.

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 healthiest period $t + 1$ health state, $\lambda_{t+1} = 1$, at the start of the next period $t + 1$].

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

The premiums for these contracts will in equilibrium be:

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

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:

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

■