

Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
<http://economics.sas.upenn.edu/pier>

PIER Working Paper 14-023

“Analyzing the Effects of Insuring Health Risks:
On the Trade-off between Short Run Insurance Benefits vs. Long Run
Incentive Costs”

by

Harold L. Cole, Soojin Kim, and Dirk Krueger

<http://ssrn.com/abstract=2430297>

Analyzing the Effects of Insuring Health Risks:*

On the Trade-off between Short Run Insurance Benefits vs. Long Run Incentive Costs

Harold L. Cole
University of Pennsylvania and NBER

Soojin Kim
Purdue University

Dirk Krueger
University of Pennsylvania, CEPR, CFS, NBER and Netspar

April 24, 2014

Abstract

This paper constructs a dynamic model of health insurance to evaluate the short- and long-run effects of policies that prevent firms from conditioning wages on health conditions of their workers, and that prevent health insurance companies from charging individuals with adverse health conditions higher insurance premia. Our study is motivated by recent US legislation that has tightened regulations on wage discrimination against workers with poorer health status (such as Americans with Disability Act of 1990, ADA, and its amendment in 2008, the ADAAA) and that prohibits health insurance companies from charging different premiums for workers of different health status starting in 2014 (Patient Protection and Affordable Care Act, PPACA). In the model, a trade-off arises between the static gains from better insurance against poor health induced by these policies and their adverse dynamic incentive effects on household efforts to lead a healthy life. Using household panel data from the PSID we estimate and calibrate the model and then use it to evaluate the static and dynamic consequences of no-wage discrimination and no-prior conditions laws for the evolution of the cross-sectional health and consumption distribution of a cohort of households, as well as ex-ante lifetime utility of a typical member of this cohort. In our quantitative analysis we find that although the competitive equilibrium features too little consumption insurance and a combination of both policies is effective in providing such insurance period by period, it is suboptimal to introduce both policies *jointly* since such a policy innovation severely undermines the incentives to lead healthier lives and thus induces a more rapid deterioration of the cohort health distribution over time. This effect more than offsets the static gains from better consumption insurance so that expected discounted lifetime utility is lower under both policies, relative to implementing wage nondiscrimination legislation *alone*. This is true despite the fact that both policy options are strongly welfare improving relative to the competitive equilibrium.

JEL Codes: E61, H31, I18

Keywords: Health Risks, Social Insurance, Health Effort Choices

*We thank Flavio Cunha, Hanming Fang, Roozbeh Hosseini, Iouri Manovskii, Ken Wolpin and participants at ASU, CUNY, EUI, EIEF, Princeton, the Board of Governors, the BYU Public Economics Conference, the Macroeconomics and Survey Data Conference in Munich, the Penn Macro Lunch and Macro Seminar, as well as the Annual Meeting for Society for Economic Dynamics 2012 and the 2013 NBER Summer Institute for many helpful comments. Krueger gratefully acknowledges financial support from the NSF under grant SES-0820494.

1 Introduction

In this paper we study the impact of social insurance policies aimed at reducing households' exposure to health related risk during their working life. These risks come through higher medical costs, higher medical premia and lower earnings. Historically there have been major insurance efforts aimed at the elderly through Medicaid and Social Security, and the poor through Medicare and income support programs like Welfare and Food Stamps. Recently the extent of these programs and the scope of the different groups they impact on has been greatly expanded. On the health insurance front, the Health Insurance Portability and Accountability Act (HIPAA) in 1996 and the Patient Protection and Affordable Care Act (PPACA) in 2010 sought to increase access to health care and to prevent health insurance being differentially priced based upon pre-existing conditions. On the income front, the 1990 Americans with Disabilities Act and its Amendment in 2008 sought to restrict the ability of employers to employ and compensate workers differentially based upon health related reasons.

In order to analyze the impact of these policies we construct a dynamic model of health insurance with heterogeneous households. As in the seminal paper by Grossman (1972), health status is an individual state variable that helps to determine both a household's productivity at work and the likelihood that she will be subject to an adverse health-related productivity shock. These in turn can be offset by medical expenditures¹ (as in Dey and Flinn 2005). The individual health status itself is persistent and changes stochastically over time, and its evolution is affected by the household's efforts to maintain their health, such as exercising and abstention from smoking. This stochastic link between effort and future health status in the model results in a moral hazard problem as health-related social insurance policies reduce households' incentives to maintain their health. We explicitly model the choice of medical expenditures and thereby endogenously determine private health insurance policies, as a function of the individual state variables of the household (health status, age and education). As an important consequence, private insurance contracts endogenously respond to the public policy regime in place.

The main question we ask is how the distributions of health status, earnings and health insurance costs will evolve under different public policy choices, and, eventually, what the impact of these choices is on social welfare. First, we study a competitive equilibrium in which workers enter into one-period employment and insurance contracts. Competition leads these contracts to partially insure the worker against within-period temporary health shocks, but not against his initial health status and its transition over time. This is also the version of the model we use for estimation and calibration to the data. In order to highlight the sources of inefficiency of the equilibrium allocation we contrast it with the choices of a constrained social planner that can dictate both the health expenditure allocation and the extent of consumption transfers across health types, but has to respect the dynamically optimal effort choices of households given these transfers. We then introduce the wage-based and the health-insurance based social insurance policies into the competitive framework. The first policy is a no-prior conditions restriction on health insurance in which competing health insurance companies cannot differentially charge premia based upon a worker's health status. The second policy is a version of a wage no-discrimination law in which firms cannot differentially hire or pay workers based upon their health status. Finally, we study the impact of implementing both policies jointly.

We study both the static and the dynamic impact of these policies. The static analysis holds the population health distribution fixed and focuses on the equilibrium health insurance contract and the provision of consumption insurance against adverse health status by the policies. In contrast, the key aspect of the dynamic analysis is the impact the policies have on individuals' incentives to maintain their health and the interaction this creates between the health distribution of the population and the costs of health insurance and productivity of the workforce. Consistent with this focus, the theoretical sections 4 and 5 focus on the productive consequences of discretionary medical expenditures.²

¹We also model catastrophic health shocks which require nondiscretionary health expenditures to avoid death.

²When we take the model to data, we also include (in our interpretation) nondiscretionary medical expenditures for major (life threatening) health events like heart attacks.

As another important simplification, while we discuss augmenting the model to include the uninsured individuals in our sensitivity analysis, we neglect them in the main analysis. With respect to discretionary spending aimed at improving productivity, the natural interpretation of the insured is that for these individuals the costs outweigh the productivity benefits and that it is the "nonproductive" gains from health expenditures that drive the decision to extend insurance to the uninsured. Since these are more difficult to quantify we don't attempt to do so in our main analysis.

To empirically implement our quantitative analysis we first estimate and calibrate the model using PSID data to match key statistics on labor earnings, medical expenditures and observed physical exercise levels. We then use the parameterized version of the model as a laboratory to evaluate the consequences of the different policy options. Our results show that a combination of wage non-discrimination law and no prior conditions law provides full insurance against health risks and implements the first-best consumption insurance allocation in the short run (that is, in the static model), but leads to a severe deterioration of incentives and thus the population health distribution in the long run (that is, in the dynamic model). This effect more than offsets the static gains from better consumption insurance so that expected discounted lifetime utility is lower under both policies, relative to implementing wage nondiscrimination legislation *alone*. Both policy options are strongly welfare improving relative to the competitive equilibrium, however.

The paper is organized as follows. After briefly reviewing the related literature in section 2 we discuss the actual U.S. policies that motivate our study. We describe the model and implementation of the two policies in the context of the model in section 3. The theoretical analysis of the static and dynamic version of the baseline model is contained in sections 4 and 5, correspondingly. In section 6 we describe how we augment the model to map it into the data, as well as our estimation and calibration procedure. Section 7 contains the main quantitative results of the paper and 8 concludes. Proofs of propositions as well as further extensions and sensitivity analyses are relegated to the appendix.

1.1 Related Literature

This paper most closely contributes to the broad literature that examines the macroeconomic and distributional implications of health, health insurance and health care policy reform. Important related contributions include Grossman (1972), Ehrlich and Becker (1972), Ehrlich and Chuma (1990), French and Jones (2004), Hall and Jones (2007), Jeske and Kitao (2009), Jung and Tran (2010), Attanasio, Kitao and Violante (2011), Ales, Hosseini and Jones (2012), Halliday, He and Zhang (2012), Hansen, Hsu and Lee (2012), Kopecky and Koleshikova (2012), Laun (2012) and Ozkan (2012) as well as Pashchenko and Porapakkam (2012).³ From this literature, perhaps most related to our work are the papers by Brügemann and Manovskii (2010) and Jung and Tran (2014) who also study the effects of PPACA on health insurance coverage and macroeconomic aggregates, but do not focus on the incentive effects of the regulation in both the labor as well as the health insurance markets (and crucially, their interaction) that we formalize in our model.

Our model incorporates health as a productive factor into a life cycle model, and studies the effect of labor- and health insurance market policies on the evolution of the cross-sectional health distribution. The dynamics of health transitions (but not current health status) in our model is affected by costly effort choices, which empirically we proxy by physical exercise and the extent of smoking. Smoking is well known to negatively affect health over time. Colman and Dave (2013) provide empirical support for the importance of exercise when they find that, controlling for observables and accounting for unobserved heterogeneity, physical activity reduces the risk of heart disease, and that the effect of *past* physical activity has a larger impact on current health status than current physical activity.

In the model health shocks and earnings interact through worker productivity. We model explicitly medical expenditures which mitigate the impact of these health shocks. A number of studies empirically estimates the effect of health on wages. These papers (see the summary in Currie and Madrian, 1999) generally find that poor health decreases wages, both directly and indirectly through a decrease in hours worked. The effect of a health shock on wages ranges from 1% to as high as 15%. Many studies consistently find that the effects on hours worked is greater than that on wages. Specifically relevant for us is Cawley's (2004) study of the impact of obesity on wages.⁴

Similarly to what we do for working age individuals, Pijoan-Mas and Rios-Rull (2012), use HRS data on self-reported health status to estimate a health transition function from age 50 onwards. They find that there

³Related, a recent literature examines the impact of health on savings and portfolio choice in life cycle models that share elements with our framework. These include Yogo (2009), Edwards (2008) and Hugonnier et al. (2012). The latter study jointly portfolio of health and other asset choices. In their model health increases productivity (labor income) and decreases occurrence of morbidity and mortality shock arrival rates (as they do in our model). The paper argues that in order to explain the correlation between financial and health status, these should be modeled jointly.

⁴In addition, there is a substantial empirical literature that estimates the effects of health inputs (and primarily, nutrition) on health outcomes and wages in developing countries. Weil (2007) surveys this literature and uses its findings to determine the impact of country-wide health differences on cross-country income variation.

is an important dependence in this transition function on socioeconomic status (most importantly education), and that this dependence is quantitatively crucial for explaining longevity differentials by socioeconomic groups. Hai (2012) and Prados (2012) also model the interaction between health and earnings over the life cycle, but their (microeconomic) focus is on wage-, earnings- and health insurance inequality.⁵

A relatively small but growing literature examines the incentive linkages between the provision of health insurance and the distribution health status in the population. Bhattacharya et al. (2009) use evidence from a Rand health insurance experiment, which featured randomized assignment to health insurance contracts, to show that access to health insurance leads to increases in body mass and obesity. They argue that this comes from the fact that insurance, especially through its pooling effect, insulates people from the impact of their excess weight on their medical expenditure costs. Consistent with this, they find the impact of being health-insured is larger for public insurance programs than in private ones in which the health insurance premium is more likely to reflect the individuals' body mass.⁶

On the policy side, several papers investigate the impact of regulation designed to limit the direct effect of health on both health insurance costs and on wages.⁷ Short and Lair (1994) examine how health status interacts with insurance choices. Madrian (1994) studies the lock-in effect of employer provided health care. On the theoretical side, Dey and Flinn (2005) estimate a model of health insurance with search, matching and bargaining and argue that employer provided health care insurance leads to reasonably efficient outcomes. Aizawa and Fang (2013) estimates a similar model with Dey and Flinn, but incorporates health and endogenize health insurance premium in order to conduct the policy effects of the ACA. Neither paper is concerned with the interaction of social insurance policies in the labor and health sector and its combined effects on household health effort choice.

Finally, related to our study of wage non-discrimination laws is the literature that studies the effect of the ADA legislation of 1990 on employment, wages and labor hours of the disabled (see DeLeire (2001) and Acemoglu and Angrist (2001), for example). Most find that it has decreased the employment of the disabled. DeLeire (2001) quantifies the effect of ADA on wages of disabled workers and reports that the negative effect of poor health on the earnings of the disabled fell by 11.3% due to ADA.

2 Institutional Background

Welfare programs in the U.S., which date back to the 1930s and were greatly expanded by the Great Society in the 1960s, insure workers against a variety of shocks, implicitly including health related shocks if they affect earnings. Since 1965 Medicare has sought to provide health insurance to the elderly and the disabled. Medicaid has sought to provide health insurance to the poor since the 1990s. Crucially for our paper, in the last two decades legislation in the U.S. was passed that limits the ability of employers to condition wages on the health conditions of employees, and to discriminate against applicants with prior health conditions when filling vacant positions.

2.1 Wage and Employment Based Discrimination

In 1990 Congress enacted the Americans with Disabilities Act (ADA) to ensure that the disabled have equal access to employment opportunities.⁸ The ADA interprets a disability as an impairment that prevents or severely restricts an individual from performing activities that are of central importance to one's daily life. Before the ADA job seekers could be asked about their medical conditions and were often required to submit to a medical exam. The act prohibited certain inquiries and conducting a medical exam before making an employment offer. However, the job could still be conditioned upon successful completion of a medical exam.⁹ The ADA permits an employer to establish job-related qualifications on the basis of business necessity. However, business necessity is limited to essential functions of the job. So impairments

⁵Both papers also study the impact of compulsory health insurance legislation.

⁶Kowalski (2012) empirically investigates, in the context of a static model, the trade-off between insurance and the moral hazard effects of health insurance provision on medical spending, finding that the latter outweigh the former.

⁷A mostly empirical literature studies the impact of the 2006 health care reform in Massachusetts. See e.g. Miller (2011).

⁸The ADA sets the federal minimum standard of protection. States may have a more stringent level.

⁹For example, the Equal Employment Opportunity Commission (EEOC) has ruled that an employee may be asked "how many days were you absent from work?", but not "how many days were you sick?".

that would only occasionally interfere with the employee's ability to perform tasks cannot be included on this list.¹⁰ Furthermore, a core requirement of the ADA is the obligation of the employer to make a reasonable accommodation to qualified disabled people.¹¹

The ADA Amendment Act (ADAAA) of 2008 rejected the strict interpretation of the ADA, broadening the notion of a disability. This included prohibiting the consideration of measures that reduce or mitigate the impact of a disability in determining whether someone is disabled. It also allowed people who are discriminated against on the basis of a perceived disability to pursue a claim on the basis of the ADA regardless of whether the disability limits, or is perceived to limit a major life activity.¹² Under the ADAAA people can be disabled even if their disability is episodic or in remission. For example people whose cancer is in remission or whose diabetes is controlled by medication, or whose seizures are prevented by medication, or who can function at a high level with learning disabilities are still disabled under the act.

2.2 Insurance Cost and Exclusion Discrimination

The Health Insurance Portability and Accountability Act (HIPAA) of 1996 placed limits on the extent to which insurance companies could exclude people or deny coverage based upon pre-existing conditions. Insurance companies were allowed exclusion periods for coverage of pre-existing conditions, but these exclusion periods were reduced by the extent of prior insurance.¹³ However, insurers were still allowed to charge higher premiums based upon initial conditions, limit coverage and set lifetime limits on benefits.¹⁴ There ample is evidence that many patients with pre-existing conditions ended up either being denied coverage or having their access to benefits limited.¹⁵

The Patient Protection and Affordable Care Act of 2010 extended protection against pre-existing conditions. Beginning in 2010 children below the age of 19 could not be excluded from their parents' health insurance policy or denied treatment for pre-existing conditions. Beginning in 2014 this restriction applies to adults as well. Moreover, insurance companies will no longer be able to use health status to determine eligibility, benefits or premia. In addition, insurers are prevented from limiting lifetime or annual benefits.¹⁶

2.3 Summary

It is our interpretation of the above legislative changes that, relative to 20 years ago, it is much more difficult now for employers to condition wages on the health status of their (potential) employees or to preferentially hire workers with better health. In addition, current legislation has made it increasingly difficult to condition the acceptance into, and insurance premia of health insurance plans on prior health conditions.

The purpose of the remainder of this paper is to analyze the aggregate and distributional consequences of these two legislative trends in the short and in the long run, with specific focus on their interactions.

3 The Model

Time $t = 0, 1, 2, \dots, T$ is discrete and finite and the economy is populated by a cohort of a continuum of individuals of mass 1. Since we are modeling a given cohort of individuals we will use time and the age of households interchangeably. We think of T as the end of working life of the age cohort under study.

¹⁰For example, an employer cannot require a driver's license for a clerking job because it would occasionally be useful to have that employee run errands. Also qualification cannot be such that a reasonable accommodation would allow the employee to perform the task. A job function is essential if the job exists to perform that function, or if the limited number of employees available at the firm requires that the task must be performed by this worker.

¹¹These accommodations include: a) making existing facilities accessible and usable b) job restructuring c) part-time or modified work schedules d) reassigning a disabled employee to a vacant position e) acquiring or modifying equipment or devices f) providing qualified readers or interpreters.

¹²Under the ADAAA major life activities now include: caring for oneself, performing manual tasks, seeing, hearing, eating, sleeping, walking, standing, lifting, bending, speaking, breathing, learning, reading, concentrating, thinking, communicating, working, as well as major bodily functions.

¹³In particular, if an individual had at least a full year of prior health insurance and she enrolled in a new plan with a break of less than 63 days, she could not be denied coverage.

¹⁴See <http://www.healthcare.gov/center/reports/preexisting.html>

¹⁵See Kass et al. (2007) and Sommers (2006).

¹⁶See again <http://www.healthcare.gov/center/reports/preexisting.html>

3.1 Endowments and Preferences

Households are endowed with one unit of time which they supply inelastically to the market. They are also endowed with an initial level of health h and we denote by $H = \{h_1, \dots, h_N\}$ the finite set of possible health levels. Households value current consumption c and dislike the effort e that helps maintain their health. We will assume that their preferences are additively separable over time, and that they discount the future at time discount factor β . We will also assume that preferences are separable between consumption and effort, and that households value consumption according to the common period utility function $u(c)$ and value effort according to the period disutility function $q(e)$.

We will denote the probability distribution over the health status h at the beginning of period t by $\Phi_t(h)$, and denote by $\Phi_0(h)$ the initial distribution over this characteristic.

Assumption 1 *The utility function u is twice differentiable, strictly increasing and strictly concave. q is twice differentiable, strictly increasing, strictly convex, with $q(0) = q'(0) = 0$ and $\lim_{e \rightarrow \infty} q'(e) = \infty$.*

3.2 Technology

3.2.1 Health Technology

Let ε denote the current health shock.¹⁷ In every period households with current health h remain healthy (that is, $\varepsilon = 0$) with probability $g(h)$. With probability $1 - g(h)$ the household draws a health shock $\varepsilon \in (0, \bar{\varepsilon}]$ which is distributed according to the probability density function $f(\varepsilon)$.

Assumption 2 *f is continuous and g is twice differentiable with $g(h) \in [0, 1]$, and $g'(h) > 0, g''(h) < 0$ for all $h \in H$.*

An individual's health status evolves stochastically over time, according to the Markov transition function $Q(h', h; e)$, where $e \geq 0$ is the level of exercise by the individual. We impose the following assumption on the Markov transition function Q

Assumption 3 *If $e' > e$ then $Q(h', h; e')$ first order stochastically dominates $Q(h', h; e)$.*

3.2.2 Production Technology

A individual with health status h and current health shock ε that consumes health expenditures x produces $F(h, \varepsilon - x)$ units of output.

Assumption 4 *F is continuously differentiable in both arguments, increasing in h , and satisfies $F(h, y) = F(h, 0)$ for all $y \leq 0$, and $F_2(h, y) < 0$ as well as $F_2(h, \bar{\varepsilon}) < -1$. Finally $F_{22}(h, y) < 0$ for all $y > 0$ and $F_{12}(h, y) \geq 0$.*

The left panel of figure 1 displays the production function $F(h, \cdot)$, for two different levels of the current health shock. Holding health status h constant, output is decreasing in the uncured portion of the health shock $\varepsilon - x$, and the decline is more rapid for lower levels of health ($h^* < h$). The right panel of figure 1 displays the production function as function of health expenditures x , for a fixed level of the shock ε , and shows that expenditures x exceeding the health shock ε leave output $F(h, \varepsilon - x)$ unaffected (and thus are suboptimal). Furthermore, a reduction of the shock ε to a lower level, ε^* , shifts the point at which health expenditures x become ineffective to the left.

The assumptions on the production function F imply that health expenditures can offset the impact of a health shock on productivity, but not raise an individual's productivity above what it would be if there had been no shock. In addition, the last assumption on F that $F_{12} \geq 0$ implies that the negative impact of a given net health shock y is lower the healthier a person is.¹⁸ The assumption $F_2(h, \bar{\varepsilon}) < -1$ insures that, if hit by the worst health shock the cost of treating this health shock, at the margin, is smaller than the positive impact on productivity (output) this treatment has.

¹⁷In the quantitative analysis we will introduce a second, fully insured (by assumption) health shock to provide a more accurate map between our model and the health expenditure data.

¹⁸This assumption is intuitive and accords with the data, as we later discuss. This is also the approach taken by Hugonnet et al. (2012) and Ehrlich and Chuma (1990).

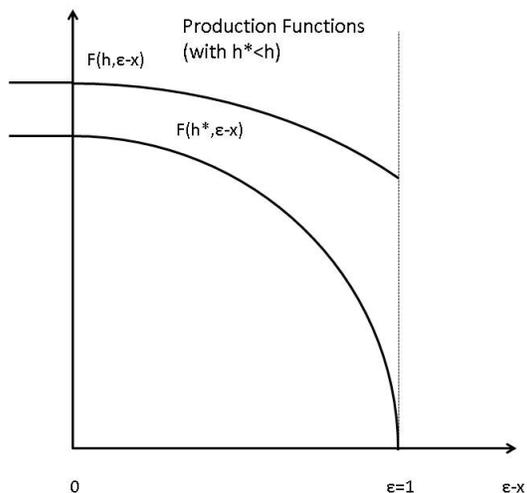


Figure 1: Production Function $F(h, \varepsilon - x)$

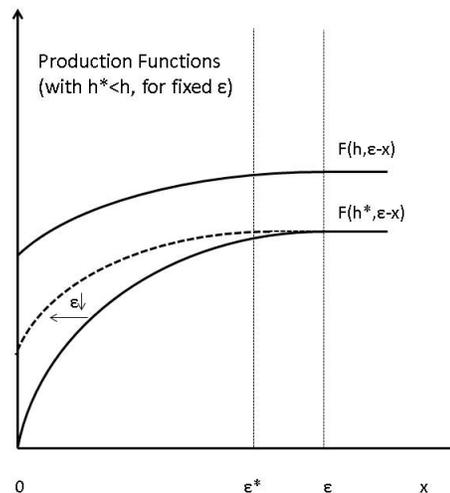


Figure 2: Production Function $F(h, \varepsilon - x)$ for fixed ε

3.3 Time Line of Events

In the current period the timing of events is as follows

1. Households enter the period with current health status h .
2. Firms offer wage $w(h)$ and health insurance contracts $\{x(\varepsilon, h), P(h)\}$ ¹⁹ to households with health status h which these households accept.
3. The health shock ε is drawn according to the distributions g, f .
4. Resources on health $x = x(\varepsilon, h)$ are spent.
5. Production and consumption takes place.
6. Households choose e .
7. The new health status h' of a household is drawn according to the health transition function Q .

3.4 Market Structure without Government

There are a large number of production firms that in each period compete for workers. Firms observe the health status of a worker h and then, prior to the realization of the health shocks, compete for workers of type h by offering a wage $w(h)$ that pools the risk of the health shocks and bundle the wage with an associated health insurance contract (specifying health expenditures $x(\varepsilon, h)$ and an insurance premium $P(h)$) that breaks even. Perfect competition for workers of type h requires that the combined wage and health insurance contract maximize period utility of the household, subject to the firm breaking even.²⁰

In the absence of government intervention a firm specializing on workers of health type h therefore offers a wage $w^{CE}(h)$ (where CE stands for competitive equilibrium) and health insurance contract $(x^{CE}(\varepsilon, h), P^{CE}(h))$ that solves

¹⁹Since we restrict attention to static contracts, whether firm offers contracts before or after the effort is undertaken does not matter.

²⁰Note that instead of assuming that firms completely specialize by hiring only a specific health type of workers h we could alternatively consider a market structure in which all firms are representative in terms of hiring workers of health types according to the population distribution and pay workers of different health h differential wages according to the schedule $w^{CE}(h)$. In other words health variation in wages and variation in hired health types h are perfect substitutes at the level of the individual firm in terms of supporting the competitive equilibrium allocation.

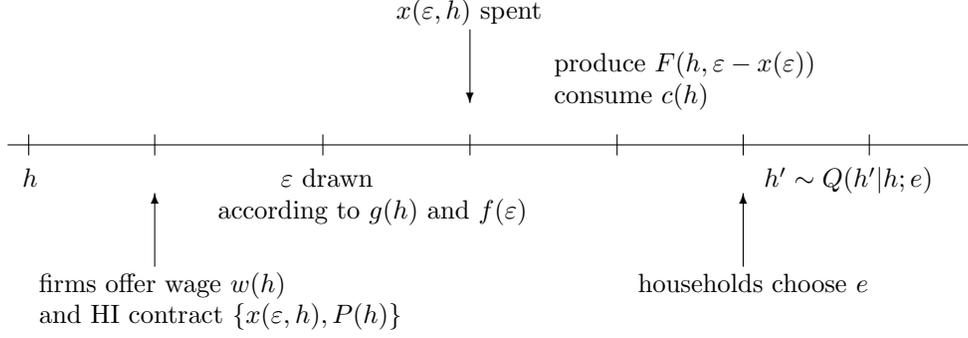


Figure 3: Timing of the Model

$$U^{CE}(h) = \max_{w(h), x(\varepsilon, h), P(h)} u(w(h) - P(h)) \quad (1)$$

s.t.

$$P(h) = g(h)x(0, h) + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)x(\varepsilon, h)d\varepsilon \quad (2)$$

$$w(h) = g(h)F(h, -x(0, h)) + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)F(h, \varepsilon - x(\varepsilon, h))d\varepsilon \quad (3)$$

Note that by bundling wages and health insurance the firm provides efficient insurance against health shocks ε , and the only source of risk remaining in the competitive equilibrium is health status risk associated with h . This risk stems both from the dependence of wages $w(h)$ as well as health insurance premia $P(h)$ on h in the competitive equilibrium, and these are exactly the sources of consumption risk that government policies preventing wage discrimination and prohibiting prior health conditions to affect insurance premia are designed to tackle.

3.5 Government Policies

We now describe how we operationalize policies that outlaws health insurance premia to be conditioned on prior health conditions h , and that limits the extent to which firms can pay workers of varying health h differential wages.

3.5.1 No Prior Conditions Law

Under this law health insurance companies are assumed to be constrained in terms of their pricing, their insurance schedule offers and their applicant acceptance criteria. The purpose of these constraints is to prevent the companies from differentially pricing insurance based upon health status.²¹ To be completely successful, these constraints must lead to a pooling equilibrium in which all individuals are insured at the same price. The best such regulation in addition assures that the equilibrium health insurance schedule $x(\varepsilon, h)$, given the constraints, is efficient. We now describe the regulations sufficient to achieve this goal.

The first constraint on health insurers is that a company must specify the total number of contracts that it wishes to issue, it must charge a fixed price independent of health status, and accept applications in

²¹Consistent with this restricted purpose, we will assume that the government cannot use health insurance to offset underlying differences in productivity coming from, say, education. This will prove important in the quantitative section.

their order of application up to the sales limit of the company. In this way, the insurance company cannot examine applications first and then decide whether or not to offer the applicant a health insurance contract.

The second constraint regulates the health expenditure schedule. If the no-prior conditions law is to have any bite the government needs to prevent the emergence of a separating equilibrium in which the health insurance companies (or the production firms in case they offer health insurance contracts) use the health expenditure schedule $x(\varepsilon, h)$ to effectively select the desired health types, given that they are barred from conditioning the health insurance premium P on h directly. Therefore, to achieve any sort of pooling in the health insurance market requires the government to regulate the health expenditure schedule $x(\varepsilon, h)$. To give the legislation the best chance of being successful we will assume that the government *regulates the health expenditure schedule $x(\varepsilon, h)$ efficiently*. For the same reason, since risk pooling is limited if some household types h choose not to buy insurance, we assume that all individuals are *forced* to buy insurance.

Given this structure of regulation and a cross-sectional distribution of workers by health type, Φ , the health insurance premium P charged by competitive firms (or competitive insurance companies, who offer health insurance in the model), given the set of regulations spelled out above, is determined by

$$P = \sum_h \left[g(h)x(0, h) + (1 - g(h)) \int f(\varepsilon)x(\varepsilon, h)d\varepsilon \right] \Phi(h) \quad (4)$$

where $x(\varepsilon, h)$ is the expenditure schedule regulated by the government. This schedule is chosen to maximize

$$\sum_h u(w(h) - P)\Phi(h)$$

with wages $w(h)$ determined by (3).

3.5.2 No Wage Discrimination Law

The objective of the government is to prevent workers with a lower health status h , and hence lower productivity, being paid less. As with the no prior conditions law, the purpose of this legislation is to help insure workers against their health status risk. However, if a production firm is penalized for paying workers with low health status h low wages, but not for preferentially hiring workers with a favorable health status (high h), then a firm can effectively circumvent the wage nondiscrimination law. Therefore, to be effective such a law must penalize *both* wage discrimination and hiring discrimination by health status.

Limiting wage dispersion with respect to gross wages $w(h)$ via legislation necessitates regulation of the health insurance market as well, in order to prevent the insurance gains from decreasing wage dispersion being undone through the adjustment of employer-provided health insurance. For example, the firm could also offer health insurance and overcharge low productivity workers and undercharge high productivity workers for this insurance, effectively undermining the illegal wage discrimination. This suggests that the government will need to limit the extent to which the cost of a worker's health insurance contracts deviates from its actuarially fair value. However, this will not be sufficient to make this policy effective.

Since the productivity of a worker depends upon the extent of his health insurance, workers whose expected productivity is below their wage will face pressure to increase their productivity through increased spending on health (and hence better health insurance coverage) while those whose productivity is above their wage will have an incentive to lower their health insurance purchases. To prevent these distortions in the health insurance market and thereby achieve better consumption insurance across h types, policy makers will need to regulate the health insurance directly as well. The moderate version of health insurance regulation would be to ensure that each policy was individually optimal and actuarially fair. The most extreme version of regulation would be to combine no-wage discrimination legislation with no-prior conditions legislation and thereby achieve the static first-best, full insurance outcome. In this case health insurance would be socially efficient and actuarially fair on average (that is, across the insured population).

We will analyze both cases. It will turn out that limiting wage dispersion with respect to net wages, $w(h) - P(h)$, avoids the negative incentive effects on the health insurance market. The policy of combining both no-wage discrimination and no-prior conditions can therefore be implemented through a policy of limiting net wage dispersion. The impact of the nondiscrimination law will, unfortunately, be sensitive to

the way in which the law is implemented, and in particular, to the form of punishment used. If the limitation in wage variation is achieved through a policy that penalizes the firms for discriminating, then these costs are realized in equilibrium, reducing overall efficiency in the economy. If, however, the limitation on wage variation is achieved either through the threat of punishment (e.g. through grim trigger strategies in repeated interactions between firms and the government) or through the delegation of hiring in a union hiring hall type arrangement, then costs from the wage nondiscrimination law will not be realized in equilibrium.²²

Since we wish to give the no wage discrimination law the best shot of being successful, in the main text we focus on the version of the policy in which no costs from the policy are realized in equilibrium, leaving the analysis of the alternative case to appendix B.2 and B.3, where we show that these policies have very negative consequences under a realized punishment regime. In either case we only tackle the extreme versions of these policies in which there is no wage discrimination (rather than limited wage discrimination) in equilibrium for reasons of analytic tractability. Under the policy, the firm takes as given thresholds on the size of the gap in wages or employment shares that will trigger the punishment. Assume that the wage penalty will be imposed if the maximum wage gap within the firm exceeds the threshold ε_w . Since type $h = 0$ will receive the lowest wage in equilibrium, to avoid the penalty a firm has to offer a wage schedule that satisfies:

$$\max_h |w(h) - w(0)| \leq \varepsilon_w.$$

Letting $n(h)$ denote the number of workers of type h hired by the firm, assume that the penalty will be imposed if the employment share of type h deviates from the population average by more than δ , and hence

$$\left| \frac{n(h)}{\sum_h n(h)} - \frac{\Phi(h)}{\sum_h \Phi(h)} \right| \leq \delta.$$

We will assume that the punishment is sufficiently dire that the firm will never choose to violate these thresholds.

We analyze the more general case in appendix B.1, but here focus on the limiting case in which the thresholds ε_w and δ converge to zero. In this case, the firm will simply take as given the economy-wide wage w^* at which it can hire a representative worker. We assume that the government regulates the insurance market determining the extent of coverage by health type, $x(e, h)$, subject to the requirement that the offered health insurance contracts exactly break even, either health type by health type (in the *absence* of a no prior conditions law) or in expectation across health types (in the *presence* of the no prior conditions law).

Perfect competition drives down equilibrium profits of firms to zero with equilibrium wages given by

$$w^* = \sum_h \left\{ g(h)F(h, -x(0, h)) + (1 - g(h)) \int_0^{\varepsilon} f(\varepsilon) [F(h, \varepsilon - x(\varepsilon, h))] d\varepsilon \right\} \Phi(h) \quad (5)$$

The insurance premium charged to the household is

$$P(h) = g(h)x(0, h) + (1 - g(h)) \int f(\varepsilon)x(\varepsilon, h)d\varepsilon \quad (6)$$

in the absence of a no-prior conditions law and

$$P = \sum_h \left[g(h)x(0, h) + (1 - g(h)) \int f(\varepsilon)x(\varepsilon, h)d\varepsilon \right] \Phi(h) \quad (7)$$

in its presence. Household consumption is given by

$$\begin{aligned} c(h) &= w^* - P(h) \text{ or} \\ c &= w^* - P \end{aligned}$$

depending on whether a no prior conditions law is in place or not.

²²The delegation method is similar to the structure we assumed in the insurance market since insurance companies were restricted to serving their customers on a first-come-first-serve basis. This assumption to us seems more problematic in the labor market because of the idiosyncratic nature of the benefits to the worker-firm match.

Given a cross-sectional health distribution Φ the efficiently regulated insurance contract $x(\varepsilon, h)$ solves:

$$\max_x \sum_h u(w^* - P(h))\Phi(h)$$

subject to (5) and (6) if the no-prior conditions restriction is not imposed on health insurance, and subject to (7) instead of (6) if the no-prior conditions restriction is present.

We now turn to the analysis of the model, starting with a static version in which by construction the choice of effort is not distorted in equilibrium. We show that in this case the competitive equilibrium implements an efficient allocation of health expenditures, but fails to provide efficient consumption insurance against prior health conditions h . We then argue that a combination of a strict wage non-discrimination law and a no prior conditions law in addition results in efficient consumption insurance in the competitive equilibrium, restoring full efficiency of allocations in the regulated market economy.

4 Analysis of the Static Model

In the analysis of the static version of our model, we will characterize both equilibrium and efficient allocations (in the absence and presence of the nondiscrimination policies). We will start with the analysis of the unregulated equilibrium, and then, by analyzing the solution to the social planner solution, highlight the source of inefficiencies (lack of consumption insurance) of the equilibrium. Finally we show that in the short run (that is statically) the combination of both policies is ideally suited to provide full consumption insurance in the regulated market equilibrium, and thus restores full efficiency of the market outcome. The static benefits of these policies are then traded off against the adverse dynamic consequences on the health distribution, as our analysis of the dynamic model will uncover in the next section.

4.1 Competitive Equilibrium

Since the only benefit of higher effort e expended by households is *dynamic* and comes in the form of a stochastically better health distribution next period, in the static version of the model effort is trivially identically equal to zero, $e(h) = 0$. As described in section 3.4 the equilibrium wage and health insurance contract solves

$$U^{CE}(h) = \max_{w(h), x(\varepsilon, h), P(h)} u(w(h) - P(h)) \quad (8)$$

s.t.

$$P(h) = g(h)x(0, h) + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)x(\varepsilon, h)d\varepsilon \quad (9)$$

$$w(h) = g(h)F(h, -x(0, h)) + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)F(h, \varepsilon - x(\varepsilon, h))d\varepsilon \quad (10)$$

The following proposition characterizes the solution to this problem:

Proposition 5 *The unique equilibrium health insurance contract and associated consumption are given by*

$$x^{CE}(\varepsilon, h) = \max [0, \varepsilon - \bar{\varepsilon}^{CE}(h)] \quad (11)$$

$$c^{CE}(\varepsilon, h) = c^{CE}(h) = w^{CE}(h) - P^{CE}(h) \quad (12)$$

$$P^{CE}(h) = (1 - g(h)) \int_{\bar{\varepsilon}^{CE}(h)}^{\bar{\varepsilon}} f(\varepsilon) [\varepsilon - \bar{\varepsilon}^{CE}(h)] d\varepsilon \quad (13)$$

$$w^{CE}(h) = g(h)F(h, 0) + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)F(h, \varepsilon - x(\varepsilon, h))d\varepsilon \quad (14)$$

where the cutoffs satisfy

$$-F_2(h, \bar{\varepsilon}^{CE}(h)) = 1. \quad (15)$$

Proof. See Appendix ■

The intuition for the equilibrium health expenditure schedule is simple. For each health status h the household cares only about net compensation $w(h) - P(h)$. By comparing equation (9) and (10) we observe that for each ε realization the marginal cost (in terms of the consumption good) of spending an extra unit of x is 1, and the marginal benefit is $-F_2(h, \varepsilon - x(\varepsilon, h))$. The optimal health expenditure schedule equates the two as long as the resulting $x(\varepsilon, h)$ is interior and features $x(\varepsilon, h) = 0$ if for a given ε the benefit of spending the first unit falls short of the cost of 1.

Corollary 6 *The cutoff $\bar{\varepsilon}^{CE}(h)$ is increasing in h , strictly so if $F_{12}(h, y) > 0$.*

The equilibrium level of health expenditure and its implications on production is graphically presented in Figure 4. As shown in the previous proposition, optimal medical expenditures take a simple cutoff rule: small health shocks $\varepsilon < \bar{\varepsilon}^{CE}(h)$ are not treated at all, but all larger shocks are fully treated up to the threshold $\bar{\varepsilon}^{CE}(h)$. These medical expenditures are displayed in Figure 4(b) for two different initial levels of health $h_1 < h_2$: below the h -specific threshold $\bar{\varepsilon}^{CE}(h)$ health expenditures are zero, and then rise one for one with the health shock ε . The determination of the threshold itself is displayed in Figure 4(a). It shows that under the assumption that the impact of health shocks on productivity is less severe for healthy households ($F_{12}(h, y) > 0$, reflected as a “more concave” curve for h_1 than for h_2 in Figure 4(a)), then the equilibrium features better insurance for less healthy households, in the sense of undoing more of the negative health shocks ε through medical treatment $x(\varepsilon, h)$. This is reflected in a lower threshold (more insurance) for h_1 than for h_2 , that is $\bar{\varepsilon}^{CE}(h_2) < \bar{\varepsilon}^{CE}(h_1)$. The equilibrium health expenditure policy function leads to a net of-health-treatment production function $F(h, \varepsilon - x^{SP}(\varepsilon, h))$ as shown in Figure 4(c).

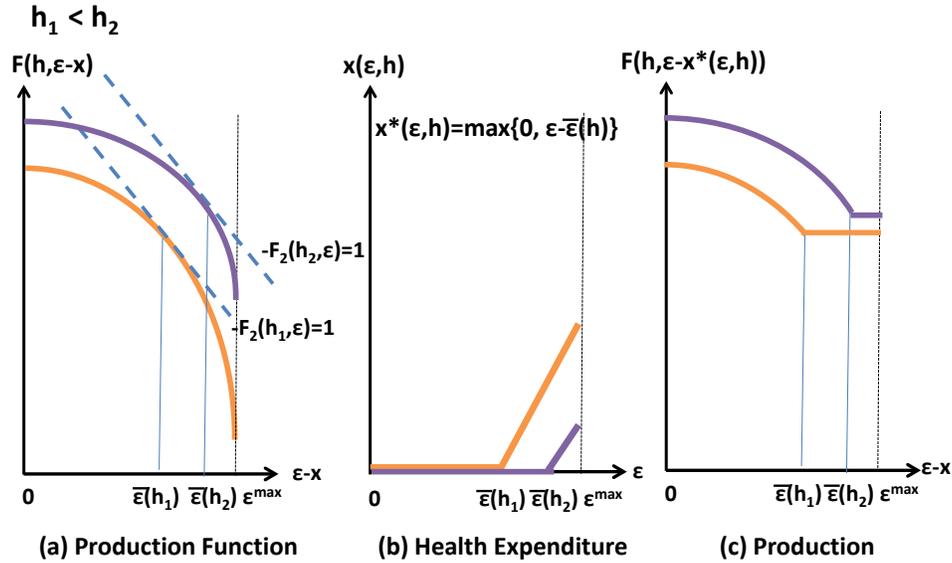


Figure 4: Optimal Health Expenditure and Production

While it follows trivially from our assumptions that the worker’s net pay, $w(h) - P(h)$, is increasing in h , it is not necessarily true that his gross wage, $w(h)$, is increasing in h as well since optimal health expenditures are decreasing in health status. We analyze the behavior of gross wages $w(h)$ with respect to health status further in Appendix C, where we provide a sufficient condition for the gross wage schedule to be monotonically increasing in h .

So far we have assumed that the production firms offer a combination of a wage and a health insurance contract to the worker, and have characterized that joint contract. We now briefly show that we can obtain the same allocation through *separate* wage contracts (offered by competitive production firms) and health insurance contracts (offered by competitive health insurers). To do that, in the next proposition we construct a wage contract offered by production firms such that it is in the worker’s interest to buy the competitive

health insurance contract characterized in proposition 5. Furthermore, the production firm's payoff under this wage contract is independent of whether a worker has in fact bought health insurance or not. As a consequence, the allocation from proposition 5 can be obtained without bundling wage and health insurance contracts, and without the need for the production firms to be able to verify whether and what health insurance workers have chosen to buy. All the firm has to observe in order to implement this wage contract is the worker's current productivity.

Proposition 7 *A wage contract of the form*

$$\mathbf{w}(h, \varepsilon - x) = \begin{cases} w^{CE}(h) & \text{if } F(h, \varepsilon - x, 0) \geq F(h, \bar{\varepsilon}^{CE}(h)) \\ w^{CE}(h) - [F(h, \bar{\varepsilon}^{CE}(h)) - F(h, \varepsilon - x)] & \text{if } F(h, \varepsilon - x, 0) < F(h, \bar{\varepsilon}^{CE}(h)) \end{cases} \quad (16)$$

offered by production firms implements the allocation characterized in proposition 5 as a competitive equilibrium in which households purchase health insurance contracts of the form given by equations 11, 13 and 15 from competitive health insurers.

Proof. The key step is to prove that given the proposed wage function the household finds it optimal to purchase the health insurance contract stipulated in proposition 5. We show this in Appendix A. ■

Note that in this decentralization the production firm implements partial consumption insurance against the ε health shocks (those below $\bar{\varepsilon}^{CE}(h)$), with the health insurance contract providing the remainder.²³.

4.2 Social Planner Problem

In order to isolate the potential inefficiencies of the competitive equilibrium allocations in the static model we now turn to an analysis of efficient allocations by characterizing the solution to the social planner problem. Given an initial cross-sectional distribution over health status in the population $\Phi(h)$ the social planner maximizes utilitarian social welfare. The social planner problem is therefore given by:

$$U^{SP}(\Phi) = \max_{x(\varepsilon, h), c(\varepsilon, h) \geq 0} \sum_h \left\{ g(h)u(c(0, h)) + (1 - g(h)) \int f(\varepsilon)u(c(\varepsilon, h))d\varepsilon \right\} \Phi(h)$$

subject to

$$\begin{aligned} & \sum_h \left\{ g(h)c(0, h) + (1 - g(h)) \int f(\varepsilon)c(\varepsilon, h)d\varepsilon + g(h)x(0, h) + (1 - g(h)) \int f(\varepsilon)x(\varepsilon, h)d\varepsilon \right\} \Phi(h) \\ & \leq \sum_h \left\{ g(h)F(h, -x(0, h)) + (1 - g(h)) \int f(\varepsilon)F(h, \varepsilon - x(\varepsilon, h))d\varepsilon \right\} \Phi(h) \end{aligned}$$

We summarize the optimal solution to the static social planner problem in the following proposition, whose proof follows directly from the first order conditions and assumption 4 (see Appendix A).

Proposition 8 *The solution to the social planner problem $\{c^{SP}(\varepsilon, h), x^{SP}(\varepsilon, h), e^{SP}(h)\}_{h \in H}$ is given by*

$$\begin{aligned} e^{SP}(h) &= 0 \\ c^{SP}(\varepsilon, h) &= c^{SP} \\ x^{SP}(\varepsilon, h) &= \max[0, \varepsilon - \bar{\varepsilon}^{SP}(h)] \end{aligned}$$

where the cutoffs satisfy

$$-F_2(h, \bar{\varepsilon}^{SP}(h)) = 1, \quad (17)$$

and the first best consumption level is given by

$$c^{SP} = \sum_h \left[g(h)F(h, 0) + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon) [F(h, \varepsilon - x^{SP}(\varepsilon, h)) - x^{SP}(\varepsilon, h)] d\varepsilon \right] \Phi(h) \quad (18)$$

The optimal cutoff $\{\bar{\varepsilon}^{SP}(h)\}$ is increasing in h , strictly so if $F_{12}(h, y) > 0$.

²³One natural interpretation of this wage contract is that the firm has limited commitment and thus can only provide partial ε insurance

The social planner finds it optimal to not have the household exercise (given that there are no dynamic benefits from doing so in the static model) and to provide full consumption insurance against adverse health shocks ε , but also against bad prior health conditions as consumption is constant in h . It follows directly from propositions 5 and 8 that

Corollary 9 *The competitive equilibrium implements the socially efficient health expenditure allocation since $\bar{c}^{CE}(h) = \bar{c}^{SP}(h)$ for all $h \in H$.*

This corollary shows that in the static case the *only* source of inefficiency of the competitive equilibrium comes from the inefficient lack of consumption insurance against adverse prior health conditions h . This can be seen by noting that

$$\begin{aligned} c^{SP} &= \sum_h \left\{ g(h)F(h, 0) + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon) [F(h, \varepsilon - x^{SP}(\varepsilon, h)) - x^{SP}(\varepsilon, h)] d\varepsilon \right\} \Phi(h) \\ &= \sum_h [w^{CE}(h) - P^{CE}(h)] \Phi(h) = \sum_h c^{CE}(h) \Phi(h) \end{aligned}$$

and thus aggregate consumption is the same in the equilibrium and efficient allocations.

In contrast to what will be the case in the dynamic model, effort trivially is not distorted in the equilibrium, relative to the allocation the social planner implements (since in both cases $e^{SP} = e^{CE} = 0$). Furthermore the equilibrium allocation of health expenditures is efficient, due to the fact that the firm bundles the determination of wages and the provision of health insurance, and thus internalizes the positive effects of health spending $x(\varepsilon, h)$ on worker productivity. However, the social planner implements complete consumption insurance against health status, $c^{SP}(h) = c^{SP}$, whereas competitive equilibrium consumption (net pay) allocations $c^{CE}(h)$ are strictly increasing in health status.

Given these results it is plausible to expect, within the context of the static model, that policies preventing competitive equilibrium wages $w^{CE}(h)$ to depend on health status (a wage non-discrimination law) and insurance premia $P^{CE}(h)$ to depend on health status (a no prior conditions law) will restore full efficiency of the policy-regulated competitive equilibrium by providing full consumption insurance. We will show next that this is indeed the case, providing a normative justification for the two policy interventions within the static version of our model.

4.3 Competitive Equilibrium with a No Prior Condition Law

As discussed above, in order to effectively implement a no prior conditions law the government has to regulate the health insurance provision done by firms or insurance companies. Given a population health distribution Φ the regulatory authority solves the problem:

$$U^{NP}(\Phi) = \max_{x(\varepsilon, h)} \sum_h u(w(h) - P) \Phi(h) \quad (19)$$

s.t.

$$P = \sum_h \left[g(h)x(0, h) + (1 - g(h)) \int f(\varepsilon)x(\varepsilon, h)d\varepsilon \right] \Phi(h) \quad (20)$$

$$w(h) = g(h)F(h, -x(0, h)) + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)F(h, \varepsilon - x(\varepsilon, h))d\varepsilon \quad (21)$$

The next proposition characterizes the resulting regulated equilibrium allocation

Proposition 10 *The equilibrium health expenditures under a no-prior condition law satisfies, for each $\tilde{h} \in H$*

$$x^{NP}(\varepsilon, \tilde{h}) = \max[0, \varepsilon - \bar{\varepsilon}^{NP}(\tilde{h})]$$

with cutoffs uniquely determined by

$$-F_2(\tilde{h}, \bar{\varepsilon}^{NP}(\tilde{h})) = \frac{\sum_h u'(w^{NP}(h) - P^{NP}) \Phi(h)}{u'(w(\tilde{h}) - P^{NP})}.$$

The equilibrium wage, for each \tilde{h} , is given by

$$w^{NP}(\tilde{h}) = g(\tilde{h})F(\tilde{h}, 0) + (1 - g(\tilde{h})) \int_0^{\bar{\varepsilon}} f(\varepsilon)[F(\tilde{h}, \varepsilon - x^{NP}(\varepsilon, \tilde{h}))]d\varepsilon$$

and the health insurance premium is determined as

$$P^{NP} = \sum_h \left[(1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)x^{NP}(\varepsilon, h)d\varepsilon \right] \Phi(h).$$

Proof. See Appendix. ■

Note that the health expenditure levels are no longer efficient as the government provides partial consumption insurance against initial health status when choosing the cutoff levels $\bar{\varepsilon}^{NP}(h)$, in the absence of direct insurance against low wages induced by bad health. In fact, as shown in the next proposition, it is efficient to over-insure households with *bad* health status and under-insure those with *good* health status, relative to the first-best.

Proposition 11 *Let \tilde{h} be the health status whose marginal utility of consumption is equal to the population average, i.e. for \tilde{h} ,*

$$-F_2(\tilde{h}, \bar{\varepsilon}(\tilde{h})) = \frac{\sum_h u'(w(h) - P)\Phi(h)}{u'(w(\tilde{h}) - P)} = 1 \quad (22)$$

holds.²⁴ Then,

$$\begin{aligned} \bar{\varepsilon}^{NP}(h) &< \bar{\varepsilon}^{SP}(h), & \text{for } h < \tilde{h} \\ \bar{\varepsilon}^{NP}(h) &= \bar{\varepsilon}^{SP}(h), & \text{for } h = \tilde{h} \\ \bar{\varepsilon}^{NP}(h) &> \bar{\varepsilon}^{SP}(h), & \text{for } h > \tilde{h}, \end{aligned}$$

The cutoffs $\bar{\varepsilon}(h)$ are strictly monotonically increasing in health status h .

Proof. See Appendix. ■

This feature of the optimal health expenditure with a no prior conditions law also indicates that mandatory participation in the health insurance contract is an important part of government regulation, since in the allocation described above healthy households cross-subsidize the unhealthy in terms of insurance premia and they are given a less generous health expenditure plan (higher thresholds) than the unhealthy.

4.4 Competitive Equilibrium with a No Wage Discrimination Law

The equilibrium with a no wage discrimination law is determined by the solution to the program:

$$\begin{aligned} U^{ND}(\Phi) &= \max_{x(\varepsilon, h)} \sum_h u(w - P(h))\Phi(h) & (23) \\ & \text{s.t.} \\ P(h) &= g(h)x(0, h) + (1 - g(h)) \int f(\varepsilon)x(\varepsilon, h)d\varepsilon \\ w &= \sum_h \left[g(h)F(h, -x(0, h)) + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)F(h, \varepsilon - x(\varepsilon, h))d\varepsilon \right] \Phi(h) \end{aligned}$$

Proposition 12 *The equilibrium health expenditures under a no-wage discrimination law alone satisfies, for each $\tilde{h} \in H$*

$$x^{ND}(\varepsilon, \tilde{h}) = \max \left[0, \varepsilon - \bar{\varepsilon}^{ND}(\tilde{h}) \right]$$

²⁴For the purpose of the proposition it does not matter whether $\tilde{h} \in H$ or not.

with cutoffs determined by

$$-F_2(\tilde{h}, \bar{\varepsilon}^{ND}(\tilde{h})) = \frac{u'(w^{ND} - P^{ND}(\tilde{h}))}{\sum_h u'(w^{ND} - P^{ND}(h))\Phi(h)}.$$

The equilibrium wage is given by

$$w^{ND} = \sum_h \left[g(h)F(h, 0) + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon) [F(h, \varepsilon - x^{ND}(\varepsilon, h))] d\varepsilon \right] \Phi(h)$$

and the health insurance premium is given by, for each \tilde{h} ,

$$P^{ND}(\tilde{h}) = (1 - g(\tilde{h})) \int_0^{\bar{\varepsilon}} f(\varepsilon)x^{ND}(\varepsilon, \tilde{h})d\varepsilon.$$

Proof. Follows directly from the first order conditions of the program (23). ■

Unlike in the no prior conditions case, we cannot establish monotonicity in the cutoffs $\bar{\varepsilon}^{ND}(\tilde{h})$. Note that under a no prior conditions law the regulatory authority partially insures consumption of the unhealthy by allocating higher medical expenditure to them. Under a no wage discrimination law instead, there are two opposing forces, preventing us from establishing monotonicity in cutoffs $\bar{\varepsilon}^{ND}(h)$ across health groups h . On one hand, a one unit increase in medical expenditure $P(h)$ is more costly to the unhealthy since marginal utility of consumption is higher for this group. On the other hand, production efficiency calls for higher medical expenditure for the unhealthy, given our assumption of $F_{12} \geq 0$ (as was the case for the no prior conditions law). Thus the cutoffs $\bar{\varepsilon}^{ND}(h)$ need not be monotone in h .²⁵

4.5 Competitive Equilibrium with Both Policies

Finally, combining both a no-wage discrimination law and a no-prior conditions legislation restores efficiency of the regulated equilibrium since both policies in conjunction provide full consumption insurance against bad health realizations h . This is the content of the next proposition.

Corollary 13 *The unique competitive equilibrium allocation in the presence of both a no wage discrimination and a no prior conditions law implements the socially efficient allocation in the static model.*

Proof. The equilibrium is the solution to

$$\begin{aligned} & \max_{x(\varepsilon, h)} \sum_h u(w^* - P)\Phi(h) \\ & \text{s.t.} \\ P &= \sum_h \left[g(h)x(0, h) + (1 - g(h)) \int f(\varepsilon)x(\varepsilon, h)d\varepsilon \right] \Phi(h) \\ w^* &= \sum_h \left[g(h)F(h, -x(0, h)) + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)F(h, \varepsilon - x(\varepsilon, h))d\varepsilon \right] \Phi(h). \end{aligned}$$

The result then follows trivially from the fact that this maximization problem is equivalent to the social planner problem analyzed above. The no prior conditions law equalizes health insurance premia P across health types, the no wage discrimination law implements a common wage w^* across health types, and the (assumed) efficient regulation of the health insurance market assures that the health expenditure schedule is efficient as well. ■

²⁵The optimal cutoff condition does imply that agent's with high marginal utility relative to the average will have lower cutoffs, and vice versa. So long as these agents still correspond to the less and more healthy agents, then we can say that the cutoffs for the less healthy have been distorted downward for the more healthy agents and upward for the less health relative to the social optimum.

Remark 14 *In appendix B.2 and B.3 we show that when the no-wage discrimination policy is enforced via punishments which are realized in equilibrium and generate real costs, then the health related productivity differences between workers are consumed by these costs and the policy has very negative outcomes. This arises because the firms must be made indifferent about who they hire at the margin in order for the labor market to clear for each health type, and this indifference is achieved by making the wage net of enforcement costs the same for all workers.*

4.6 Summary of the Analysis of the Static Model

The competitive equilibrium implements the efficient health expenditure allocation but does not insure households against initial health conditions. Both a no-prior conditions law and a no-wage discrimination law provide partial, but not complete, consumption insurance against this risk, without distorting the effort level. The health expenditure schedule is distorted when each policy is implemented in isolation, relative to the social optimum, as the government provides additional partial consumption insurance through health expenditures. Only both laws in conjunction implement a fully efficient health expenditure schedule and full consumption insurance against initial health conditions h , and therefore restore the first best allocation in the static model. Enacting both policies jointly is thus fully successful in what they are designed to achieve in a static world (partially due to the fact that additional government regulation severely restricted the options of firms to circumvent the government policies).

5 Analysis of the Dynamic Model

We now study a dynamic version of our economy. Both in terms of casting the problem, as well as in terms of its computation we make use of the fact that there is no aggregate risk (due to the continuum of agents cum law of large numbers assumption). Therefore the sequence of cross-sectional health distributions $\{\Phi_t\}_{t=0}^T$ is a deterministic sequence. Furthermore, conditional on a distribution Φ_t today the health distribution tomorrow is completely determined by the effort choice $e_t(h)$ of households²⁶ (or the social planner), so that we can write

$$\Phi_{t+1} = H(\Phi_t; e_t(\cdot)) \quad (24)$$

where the time-invariant function H is in turn completely determined by the Markov transition function $Q(h'; h, e)$. The initial distribution Φ_0 is an initial condition and exogenously given.

Under each policy, given a sequence of aggregate distributions $\{\Phi_t\}_{t=0}^T$ we can solve an appropriate dynamic maximization problem of an individual household for the sequence of optimal effort decisions $\{e_t(h)_{h \in H}\}_{t=0}^T$ which in turn imply a new sequence of aggregate distributions via (24). Our computational algorithm for solving competitive equilibria then amounts to iterating on the sequences $\{\Phi_t, e_t\}$. Within each period the timing of events follows exactly that of the static problem in the previous section.

5.1 Competitive Equilibrium without Policy

In our model, since absent wage and health insurance policies households do not interact in any way, we can solve the dynamic programming problem of each household independently of the rest of society. The only state variables of the household are her current health h and age t , and the dynamic program reads as:

$$v_t(h) = U^{CE}(h) + \max_{e_t(h)} \left\{ -q(e_t(h)) + \beta \sum_{h'} Q(h'; h, e_t(h)) v_{t+1}(h') \right\} \quad (25)$$

²⁶We assert here that the optimal effort in period t is only a function of the current individual health status h . We will discuss below the assumptions required to make this assertion correct.

where

$$\begin{aligned}
U^{CE}(h) &= \max_{x(\varepsilon, h), w(h), P(h)} u(w(h) - P(h)) \\
&\quad s.t. \\
w(h) &= g(h)F(h, -x(0, h)) + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)F(h, \varepsilon - x(\varepsilon, h))d\varepsilon \\
P(h) &= g(h)x(0, h) + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)x(\varepsilon, h)d\varepsilon
\end{aligned}$$

is the solution to the static equilibrium problem in section 4.1, which was given by:

$$\begin{aligned}
x^{CE}(\varepsilon, h) &= \max [0, \varepsilon - \bar{\varepsilon}^{CE}(h)] \\
c^{CE}(h) &= w^{CE}(h) - P^{CE}(h) \\
P^{CE}(h) &= (1 - g(h)) \int_{\bar{\varepsilon}^{CE}(h)}^{\bar{\varepsilon}} f(\varepsilon) [\varepsilon - \bar{\varepsilon}^{CE}(h)] d\varepsilon \\
w^{CE}(h) &= g(h)F(h, 0) + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)F(h, \varepsilon - x(\varepsilon, h))d\varepsilon
\end{aligned}$$

with cutoff:

$$-F_2(h, \bar{\varepsilon}^{CE}(h)) = 1$$

Note again that the provision of health insurance is socially efficient in the competitive equilibrium.

In contrast to the social planner problem, and in contrast to what will be the case in a competitive equilibrium with a no-wage discrimination law or a no-prior conditions law, in the unregulated competitive equilibrium there is no interaction between the maximization problems of individual households. Thus the dynamic household maximization problem can be solved independent of the evolution of the cross-sectional health distribution. It is a simple dynamic programming problem with terminal value function

$$v_T(h) = U^{CE}(h)$$

and can be solved by straightforward backward iteration.

Given the solution $\{e_t(h)\}$ of the household dynamic programming problem and given an initial distribution Φ_0 the dynamics of the health distribution is then determined by the aggregate law of motion (24). The optimal choice $e_t(h)$ solves the first order condition

$$q'(e_t(h)) = \beta \sum_{h'} \frac{\partial Q(h'; h, e_t(h))}{\partial e_t(h)} v_{t+1}(h') \quad (26)$$

Note that at time t when the decision $e_t(h)$ is taken the function $v_{t+1}(\cdot)$ is known. Furthermore, given knowledge of v_{t+1} and the optimal e_t the period t value function v_t is determined by (25). By assumptions 1 and 3, from equation (26) it follows that effort $e_t(h)$ is positive for all t and h .

5.2 Constrained Social Planner Problem

The purpose of studying the social planner problem is to provide a benchmark against which to compare allocations emerging in equilibrium, in the absence and presence of the nondiscrimination policies. In the static model the planner could provide consumption insurance, as could both policies. In the dynamic model with endogenous effort choice an unconstrained planner in addition can dictate effort choices, whereas both policies under consideration cannot impact effort choice directly (only indirectly, though changing the economic consequences of worse health outcomes). Thus we believe it is more instructive to study a restricted dynamic planner problem in which the social planner has to respect the intertemporal optimality condition (26) with respect to effort private households would choose, given the age- and health-dependent consumption and effort allocation $\{c_t(h), e_t(h)\}$ chosen by the planner. We think of these constraints as emerging from the inability of the planner to directly observe household effort choices: if a certain effort

$e_t(h)$ is desired by the planner, it has to be induced by a consumption allocation that makes providing that effort individually rational, given the health-dependent consumption allocations from tomorrow onward.

Letting by $V_t(h)$ denote the expected lifetime utility for a household with current age t and health status h , given by recursively by

$$V_t(h) = u(c_t(h)) - q(e_t(h)) + \beta \sum_{h'} Q(h'; h, e_t(h)) V_{t+1}(h')$$

and with exogenously given terminal conditions $\{V_{T+1}(h') \equiv 0\}$. The social planner maximizes

$$V(\Phi_0) = \sum_h V_0(h) \Phi_0(h)$$

by choice of $\{c_t(h), e_t(h)\}$ and thus implied $\{V_t(h)\}$, subject to the resource constraints

$$\sum_h c_t(h) \Phi_t(h) \leq Y(\Phi_t)$$

and the incentive constraints on effort which equate the marginal utility cost for each household to the marginal benefit of better health from tomorrow onward for that household, given a consumption allocation chosen by the planner:

$$q'(e_t(h)) = \beta \sum_{h'} \frac{\partial Q(h'; h, e_t(h))}{\partial e_t(h)} V_{t+1}(h'). \quad (27)$$

Aggregate (net of health expenditure) resources available for consumption in every period are solely a function of the cross-sectional health distribution in that period, and are given by

$$Y(\Phi) = \sum_h \left[g(h) F(h, 0) + (1 - g(h)) \int_{\varepsilon} f(\varepsilon) [F(h, \varepsilon - x^{SP}(\varepsilon, h)) - x^{SP}(\varepsilon, h)] d\varepsilon \right] \Phi(h)$$

where $x^{SP}(\varepsilon, h)$ is the efficient health expenditure schedule from the static social planner problem characterized in section 4.2:

$$x^{SP}(\varepsilon, h) = \max [0, \varepsilon - \bar{\varepsilon}^{SP}(h)] \quad (28)$$

with cutoffs defined by²⁷

$$-F_2(h, \bar{\varepsilon}^{SP}(h)) = 1. \quad (29)$$

Finally, the evolution of the cross-sectional health distribution follows the general law of motion given in (24), or more explicitly

$$\Phi_{t+1}(h') = \sum_h Q(h'; h, e_t(h)) \Phi_t(h).$$

In section D in the appendix we discuss how a solution to this problem can be determined numerically.

Equation (27) demonstrates the basic trade-off for the constrained dynamic social planner that will also be present in our evaluation of both policies. In the static analysis we showed that it was optimal for the social planner to provide full consumption insurance. Although the constrained planner can certainly implement such an allocation, it will lead to identically zero effort in all periods, on account of equation 27. We state this in the next

Proposition 15 *Suppose the planner chooses a consumption allocation $\{c_t(h)\}$ such that*

$$c_t(h) = c_t$$

for all $h \in H$, then $e_t(h) = 0$ for all t, h .

²⁷Given that health expenditures only affect current worker productivity and thus do not interact with current effort choice, there is no gain for the social planner, for a given health distribution, to deviate from static production efficiency (as manifested in equations (28) and (29))

Proof. Is identical to the same result obtained with both policies in place (proposition 16) and given there.

■

Since the marginal cost of providing effort at $e_t(h) = 0$ is zero (assumption 1) and the benefit on health transitions and thus net production and average consumption is positive (on account of assumptions 3 and 4), starting from the full consumption, zero effort allocation a marginal increase in effort is welfare improving (since the consumption insurance losses are second order when starting at the full consumption insurance allocation), and thus $e_t^{SP}(h) > 0$. This will be in contrast to the outcome under both policies (again see proposition 16) which²⁸ features full consumption insurance and zero effort, therefore resulting in an inefficient allocation, relative to the constrained social planner solution.

5.3 Competitive Equilibrium with a No Prior Condition Law

As discussed above, we assume that the government in every period t takes as given the health distribution Φ_t and enforces the no prior condition law and regulates health insurance contracts efficiently, as in the static analysis of section 4.3. We now make explicit that the solution of the static government regulation problem (19)-(21) is a function of the cross-sectional health distribution,

$$x^{NP}(\varepsilon, \tilde{h}; \Phi_t) = \max[0, \varepsilon - \bar{\varepsilon}^{NP}(\tilde{h}; \Phi_t)] \quad (30)$$

with cutoffs for each $\tilde{h} \in H$ determined by

$$-F_2(\tilde{h}, \bar{\varepsilon}^{NP}(\tilde{h}; \Phi_t))u'(w^{NP}(\tilde{h}; \Phi_t) - P^{NP}(\Phi_t)) = \sum_h u'(w^{NP}(h; \Phi_t) - P^{NP}(\Phi_t))\Phi_t(h) := Eu'(\Phi_t) \quad (31)$$

and

$$w^{NP}(h; \Phi_t) = g(h)F(h, 0) + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)[F(h, \varepsilon - x^{NP}(\varepsilon, h; \Phi_t))]d\varepsilon \quad (32)$$

$$P^{NP}(\Phi_t) = \sum_h \left[g(h)x^{NP}(0, h; \Phi_t) + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)x^{NP}(\varepsilon, h; \Phi) d\varepsilon \right] \Phi_t(h) \quad (33)$$

In order for the household to solve her dynamic programming problem she only needs to know the sequence of wages and health insurance premia $\{w_t(h), P_t\}$, but not necessarily the sequence of distributions that led to it. Given such a sequence the dynamic programming problem of the household then reads as

$$v_t(h) = u(w_t(h) - P_t) + \max_{e_t(h)} \left\{ -q(e_t(h)) + \beta \sum_{h'} Q(h'; h, e_t(h))v_{t+1}(h') \right\} \quad (34)$$

with terminal condition $v_T(h) = u(w_T(h) - P_T)$. As before the optimality condition reads as

$$q'(e_t(h)) = \beta \sum_{h'} \frac{\partial Q(h'; h, e_t(h))}{\partial e_t(h)} v_{t+1}(h'). \quad (35)$$

and thus equates the marginal cost of providing effort, $q'(e)$ with the marginal benefit of an improved health distribution tomorrow. Although equation (35) looks identical to equation (26) from the unregulated equilibrium, the determination of the value functions that appear on the right hand side of both equations is not (compare the first terms on the right hand sides of equations (25) and (34)). The difference in these equations highlights the extra consumption insurance induced by the no-prior conditions law, in that with this policy the health insurance premium does not vary with h . This extra consumption insurance, ceteris paribus, reduces the variation of v_{t+1} in h' and thus limits the incentives to exert effort in order to achieve a (stochastically) higher health level tomorrow. In appendix E we describe a computational algorithm to solve the dynamic model with a no-prior conditions law.

²⁸This statement is true absent direct utility benefits from better health, as assumed in the theoretical model but relaxed in the quantitative implementation.

5.4 Competitive Equilibrium with a No Wage Discrimination Law

The main difference to the previous section is that now the static health insurance contract and premium are given by health spending

$$x^{ND}(\varepsilon, \tilde{h}; \Phi_t) = \max[0, \varepsilon - \bar{\varepsilon}^{ND}(\tilde{h}; \Phi_t)] \quad (36)$$

with cutoffs for each $\tilde{h} \in H$ determined by

$$-F_2(h, \bar{\varepsilon}^{ND}(h))Eu'_t = u'(w^{ND}(\Phi_t) - P^{ND}(h, \Phi_t)) \quad (37)$$

where

$$Eu'_t := \sum_h u'(w^{ND}(\Phi_t) - P^{ND}(h, \Phi_t))\Phi_t(h). \quad (38)$$

The equilibrium wage is given by

$$w^{ND}(\Phi_t) = \sum_h \left\{ g(h)F(h, 0) + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)[F(h, \varepsilon - x^{ND}(\varepsilon, h; \Phi_t))]d\varepsilon \right\} \Phi_t(h). \quad (39)$$

The equilibrium health insurance premium depends on whether a no prior conditions law is in place or not: Without such policy the premia are given as

$$P^{ND}(h; \Phi_t) = P^{ND}(h) = (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)x^{ND}(\varepsilon, h)d\varepsilon \quad (40)$$

whereas with both policies in place the premium is determined by²⁹

$$P^{Both}(\Phi_t) = \sum_h \left[(1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)x^{Both}(\varepsilon, h)d\varepsilon \right] \Phi_t(h) \quad (41)$$

For a given sequence of wages $\{w_t, P_t(h)\}$ the dynamic problem of the household reads as before:

$$v_t(h) = u(w_t - P_t(h)) + \max_{e_t(h)} \left\{ -q(e_t(h)) + \beta \sum_{h'} Q(h'; h, e_t(h))v_{t+1}(h') \right\} \quad (42)$$

and the terminal condition $v_T(h) = u(w_T - P_T(h))$, first order conditions and updating of the value function for this version of the model are exactly the same, mutatis mutandis, as under the previous policy. In appendix E we discuss the algorithm to solve this version of the model.

5.5 Competitive Equilibrium with Both Laws

If both policies are in place simultaneously, we can give a full analytical characterization of the equilibrium without resorting to any numerical solution procedure. We do so in the next proposition.

Proposition 16 *Suppose there is a no wage discrimination and a no prior condition law in place simultaneously. Then*

$$e_t^{Both}(h) = 0 \text{ for all } h, \text{ and all } t.$$

The provision of health insurance is socially efficient. From the initial distribution Φ_0 the health distribution in society evolves according to (24) with $e_t(h) \equiv 0$.

Corollary 17 *The consumption-effort allocation under both policies is inefficient since it leads to inefficiently low effort.*

²⁹Wages still take the form as in (39), but with $x^{Both}(\varepsilon, h)$ replacing $x^{ND}(\varepsilon, h)$. Recall from the static analysis that $x^{Both}(\varepsilon, h) = x^{SP}(\varepsilon, h)$, that is, the medical expenditure schedule is socially efficient.

The proof is by straightforward backward induction and is given in Appendix A. In the presence of both policies there are no incentives, either through wages or health insurance premia, to exert effort to lead a healthy life. Since effort is costly, households won't provide any such effort in the regulated dynamic competitive equilibrium. Thus in the absence of any direct utility benefits of better health the combination of both policies leads to a complete collapse in incentives, with the associated adverse long run consequences for the distribution of health in society.

Equipped with these theoretical results and the numerical algorithms to solve the various versions of our model we now map our model to cross-sectional health and exercise data from the PSID to quantify the effects of government regulations on the evolution of the cross-sectional health distribution, as well as aggregate production, consumption and health expenditures.

6 Bringing the Model to the Data

Prior to spelling out the details of the empirical implementation it is useful to review the basic logic of our model, in order to highlight what aspects of the data drive our theoretical assumptions and will drive our quantitative results. It will also serve to justify the extensions of the model for the empirical analysis discussed in the next subsection, and help to clarify our choices for the estimation and calibration.

First, in the data we observe a *positive correlation between wages and health* whose magnitude will inform the estimation of the production function (and thus wages) in section 6.2.3. However, since we do not wish to attribute all cross-sectional variation in wages to health h and age t , we will permit wages to depend on other determinants as well. *Second*, the data display a *negative correlation between health expenditures and health status* which will drive the estimation of $g(h)$, the probability of a productivity-reducing health shock ε in section 6.2.2. Since not all health expenditures in the data can be characterized as productivity-enhancing we augment the health expenditure process by outlays for catastrophic events, as discussed in the next section.

These two empirical observations indicate that having a good health status h is economically beneficial (since it increases wages and lowers health expenditures), as hypothesized by our model. The *third* key observation is that, controlling for current health, empirical *measures of health effort e raise, in expectation, future health status*. This indicates that individuals can partially control health, an observation we exploit in the estimation of the transition matrix Q in section 6.2.2. Fourth, we use the fact that wage differences coming through education alter the incentive to maintain health in order to determine the cost elasticity of effort. The policy experiments then quantify, based on a parameterized version of the model that is informed by the four empirical observations stated above, the deterioration in incentives for effort due to policy induced reductions in the correlations between health h and wages $w(h)$ and health expenditures borne by the household (as measured by the health insurance premia $P(h)$).

We now briefly justify our empirical methodology that combines maximum likelihood estimation and calibration. At the core of our paper are the incentive effects on individual effort of social insurance interventions. To evaluate these, we need to estimate how effort costs $q(e)$ and health transition probabilities $Q(h'; h, e)$ change with health related effort e . The cross-sectional heterogeneity in effort-induced health transitions provides crucial evidence along this dimension, and thus we aim to fully capture this heterogeneity by estimating the transition function Q using maximum likelihood on panel data. On the other hand, self-reported health status comes in the form of a very coarse grid, and moreover, explaining other sources of cross-sectional heterogeneity that affects wages and health expenditures (which is also pervasive in the data) is not our main focus. We therefore opted for a calibration approach that fits selected averages in the data to parameterize the production and preference side of the model.

6.1 Augmenting the Model

The model described so far only included the necessary elements to highlight the key static insurance-dynamic incentive trade-off we want to emphasize. However, to insure that the model can capture the significant aspects of the health, health effort and health expenditure data observed in micro data we now augment it in four aspects. We want to stress, however, that none of the qualitative results derived so far rely on the absence of these elements, which is why we abstracted from them in our theoretical analysis.

First, in the data some households have health expenditures in a given year from catastrophic illnesses that exceed their labor earnings. In the model, the only benefit of spending resources on health is to offset the negative productivity consequences of the adverse health shocks ε . Thus it is never optimal to incur health expenditures that exceed the value of a worker’s production in a given period. In order to capture these large medical expenditures in data and arrive at realistic magnitudes of health insurance premia we introduce a second health shock. This exogenous shock z stands in for a catastrophic health expenditure shock, and when households receive the z -shock, they *have to* spend z ; otherwise, they die (or equivalently, incur a prohibitively large utility cost). Households in the augmented model are assumed to either not receive any health shock, face either a z -shock, or an ε -shock, but not both. We denote by $\mu_z(h)$ the mean of the health expenditure shock z , conditional on initial health h , and by $\kappa(h)$ the probability of receiving a positive z -shock. Households that received a z -shock can still work, but at a reduced productivity $\rho < 1$ relative to healthy workers. As described in more detail in appendix F.1, the z -shock merely scales up health insurance premia by $\mu_z(h)$ and introduces additional health-related wage risk (since z -shocks come with a loss of $1 - \rho$ of labor productivity).

Second, in our model so far all variation in wages was due to health (h and $\varepsilon - x$). When bringing the model to the data we permit earnings in the model to also depend on the age t and education $educ$ of a household, and consequently specify the production function as $F(t, educ, h, \varepsilon - x)$. Given this extension we have to take a stance on how households of different education levels interact in equilibrium under each policy. Since our objective is to highlight the insurance aspect of both policies with respect to health-related consumption risks we assume that even in the presence of a wage discrimination law individuals with higher education can be paid more, and that health insurance companies can charge differential premia to individuals with heterogeneous education levels even in the presence of a no-prior conditions law.³⁰ Consistent with the introduction of skill (education) heterogeneity the initial distribution over household types is now denoted by $\Phi_0(h, educ)$ and will be determined directly from the data.

Third, for the model to have a chance of generating the observed pattern in *exercise levels* of individuals with different health status (and specifically, the fact that less healthy individuals tend to exercise less in the data) we now introduce a health-specific preference shifter for the disutility of effort. Instead of being given by $q(e)$, as in the theoretical analysis so far, the cost of exerting effort is now assumed to be given as $\gamma(h)q(e)$. Note that since $\gamma(h)$ only affects the disutility of effort which is separable from the utility of consumption, the analysis of the static model in section 4 remains completely unchanged (and so do the optimal health insurance contracts and health expenditure allocations). In the analysis of the dynamic model all expressions involving $q(\cdot)$ turn into $\gamma(h)q(\cdot)$ but the analysis is otherwise unaltered. However, because the wage benefits of health and the impact of effort on health transitions combine to make the incentive to exert effort strongest for low health levels, $\gamma(h)$ needs to be declining health status (i.e. it is less costly for healthy people to exert effort) in order for the model to match the fact that exercise is increasing in health status.

The last, and perhaps most significant departure from the theoretical model is that we now endow the household with an education and health-dependent continuation utility $v_{T+1}(educ, h)$ from retirement. The theoretical model implicitly assumed that this continuation utility was identically equal to zero, independent of the health status at retirement. The vector $v_{T+1}(educ, h)$ will be determined as part of our structural model estimation. Endowing individuals with nontrivial continuation utility at retirement avoids the counterfactual prediction of the model that effort is zero in the last period of working life, T . This assumption also introduces a direct utility benefit from better health (albeit one that materializes at retirement) and thus avoids the complete collapse of incentives to provide effort under both policies (that is, proposition 16 no longer applies).

In the rest of this section, we use the so extended version of our model to estimate parameters to match PSID data on health, expenditure and exercise in 1999. In the main body of the paper, we describe the procedure we follow in a condensed form, relegating the detailed data description and estimation procedures to the Appendix F. Once the model is parameterized and its reasonable fit of the data established, in section 7 we then use it to analyze the positive and normative short- and long-run consequences of introducing non-discrimination legislation.

³⁰In order to obtain a meaningful welfare comparison with socially optimal allocations we also solve the social planner problem separately for each ($educ$) group, therefore ruling out ex ante social insurance against the inability to attain a higher education level.

6.2 Parameter Estimation and Calibration

The determination of the model parameters proceeds in three steps. First, we fix a small subset of parameters exogenously. Second, parts of the model parameters can be estimated from the PSID data directly. These include the parameters governing the health transition function $Q(h'|h, e)$, the probabilities $(g(h), \kappa(h))$ of receiving the ε and z health shocks, as well as the productivity effect of the z -shocks given by ρ . Third, (and given the parameters obtained in step 1 and 2) the remaining parameters (mainly those governing the production function F , the ε -shock distribution $f(\varepsilon)$ and preferences) are then determined through a method of moments estimation of the model with PSID wage, health and effort data. We now describe these three steps in greater detail.

6.2.1 A Priori Chosen Parameters

First, we choose one model period to be six years, a compromise between assuring that effort has a noticeable effect on health transitions (which requires a sufficiently long time period) and reasonable sample sizes for estimation (which speaks for short time periods).³¹ We then select two preference parameters a priori. Consistent with values commonly used in the quantitative macroeconomics literature we choose a risk aversion parameter of $\sigma = 2$ and a time discount factor of $\beta = 0.96$ *per annum*.

6.2.2 Parameters Estimated Directly from the Data

In a second step we estimate part of the model parameters directly from the data, without having to rely on the equilibrium of the model. First, we obtain the initial distribution $\Phi_0(h, educ)$ directly from the data. The following additional model elements are also directly determined from the data.

Health Transition Function $Q(h'|h, e)$ The PSID includes measures of light and heavy exercise levels³² and number of cigarettes smoked in a day starting in 1999 which we use to estimate health transition functions. We denote by e^l , e^h , and s , the frequency of light and heavy exercise levels and smoking. We normalize these effort measures so that they lie between 0 and 1, and assume the following parametric form for the health transition function:

$$Q(h'; h, e^l, e^h, 1-s) = \begin{cases} (1 + \pi(h, e^l, e^h, 1-s)^{\alpha_i(h)})G(h, h'), & \text{if } h' = h + i, i \in \{1, 2\} \\ (1 + \pi(h, e^l, e^h, 1-s))G(h, h'), & \text{if } h' = h, h > 1 \text{ or } h' = h + 1, h = 1 \\ \left(\frac{1 - \sum_{h' \geq h} Q(h'; h, e^l, e^h, 1-s)}{\sum_{h' < h} G(h, h')} \right) G(h, h'), & \text{if } h' = h - 1, h > 1 \text{ or } h' = h, h = 1 \end{cases}$$

where

$$\pi(h, e^l, e^h, 1-s) = \phi(h)(\delta_l e^l + \delta_h e^h + \delta_s(1-s))^{\lambda(h)}.$$

First note that as smoking has adverse effects on health, we use $1-s$ as a measure of (not) *smoking effort*. Moreover, given that light and heavy physical exercise and smoking can have different effects on health transition, we give weights $\{\delta_l, \delta_h, \delta_s\}$ on each component of effort, with $\delta_l + \delta_h + \delta_s = 1$. We think of $\delta_l e^l + \delta_h e^h + \delta_s(1-s)$ as the composite effort level e used in the theoretical analysis of our model.

³¹This longer period length also makes our assumption that health shocks are i.i.d. conditional on current health status more natural.

³²Number of times an individual carries out light physical activity (walking, dancing, gardening, golfing, bowling, etc.) and heavy physical activity (heavy housework, aerobics, running, swimming, or bicycling).

Health Shock Probabilities $g(h)$ and $\kappa(h)$ In our model, $g(h)$ represents the probability of *not* receiving any shock, and $\kappa(h)$ is the probability of facing a z -shock. Since we assume that households do not receive both an ε -shock and a z -shock in the same period, the probability of facing an ε -shock is given by $1 - g(h) - \kappa(h)$. From PSID, we first construct the probabilities of having a z -shock and an ε -shock. We define households that have received a z -shock as those who were diagnosed with cancer, a heart attack, or a heart disease³³ and those who spent more on medical expenditures than their current income when hit with a health shock. Households with all other health shocks or those who missed work due to an illness are categorized as having received an ε -shock.

Impact ρ of a z -shock on Productivity Using the criterion for determining ε and z -shocks specified above, we use mean earnings of those with a z -shock relative to those without any health shock to directly estimate ρ .

6.2.3 Parameters Calibrated within the Model

In a final step we now use our model to find parameters governing the production function, the ε - and z -shock distribution, the preference parameters governing disutility from effort, and the terminal value function $v_{T+1}(h')$. The structure of our model allows us to calibrate the parameters in two separate steps. The first part of the estimation consists of finding parameters for the production function and distribution of health shocks, and only involves the *static* part of the model from section 4. This is the case since realized wages and health expenditures in the model are determined exclusively from the static part of the model and are independent of effort decisions and the associated health evolution in the dynamic part of the model. In a second step we then employ the dynamic part of the model to estimate the preference parameters for exercise and the terminal value of health.³⁴

Production Function and Health Status We assume the following parametric form for the production technology:

$$F(t, educ, h, \varepsilon - x) = A(t, educ)h + \frac{(\bar{\varepsilon} - (\varepsilon - x))^{\phi(a, educ)}}{h^{\xi(a, educ)}}, \quad 0 < \phi(\cdot), \xi(\cdot) < 1, A(\cdot) > 0.$$

The production function captures two effects of health on production: the direct effect (first term) and the indirect effect which induces the marginal benefit of health expenditures x to decline with better health (that is $-F_{12} < 0$). The term $A(t, educ)$ allows for heterogeneity in age and education of the effect of health on production and thus wages. Here age can take seven values, $t \in \{1, 2, \dots, 7\}$ and we classify individuals into two education groups, those that have graduated from high school and those that have not: $educ \in \{\text{less than High School, High School Grad}\}$. We also allow for differences in marginal effects of medical expenditures on production across education and two broad age groups through parameters $\phi(a, educ)$ and $\xi(a, educ)$, where $a \in \{Young, Old\}$. We define *Young* as those individuals between the ages of 24 and 41 and the rest as *Old*. This age classification divides our sample roughly in half. We represent the functions $A(t, educ)$, $\phi(a, educ)$ and $\xi(a, educ)$ by a full set of age and education dummies.

Since in the unregulated equilibrium the production of individuals (after health expenditures have been made) equals their labor earnings, we use data on labor earnings of households with different health status $\left(\frac{w(h_2)}{w(h_1)}, \frac{w(h_3)}{w(h_1)}, \frac{w(h_4)}{w(h_1)}\right)$ as well as relative average earnings of the *Young* and the *Old* to pin down the health levels $\{h_1, h_2, h_3, h_4\}$ in the model.³⁵ Moreover, since $A(t, educ)$ captures the effects of age t and education $educ$ on labor earnings we use conditional (on age and education) earnings to pin down the 14 (7×2) parameters $A(t, educ)$.

³³These three diseases lead to the most mean medical expenditures, relative to other health conditions reported in the data.

³⁴Even though we describe the parameters and calibration targets of the different model elements in separate subsections below for expositional clarity, the parameters for production function and health shock distributions are calibrated *jointly*, using the targets in these sections. Similarly, the parameters governing preferences for exercise and the marginal value of health at the terminal date are calibrated *jointly*, using the observations in both subsections.

³⁵The categories {Excellent, Very Good, Good, Fair} used in the data itself have no cardinal interpretation.

In order to determine the values of the dummies representing $\phi(\cdot)$ and $\xi(\cdot)$ we recognize that in the model they determine the expenditure cutoffs for the ε -shock, as a function of individual health status. Thus we use medical expenditure data to estimate these parameters. More specifically the four parameters representing $\phi(a, educ)$ are determined to fit the percentage of labor earnings spent on medical expenditure (averaged over h) for each $(a, educ)$ -group and the four parameters representing $\xi(a, educ)$ are chosen to match the percentage of labor earnings spent on medical expenditures (averaged over $(a, educ)$ groups) for each level $h \in H$ of household health.³⁶

Distribution of Health Shocks In order to estimate the parameters governing the distribution of health shocks ε we exploit the theoretical result from section 4 that medical expenditures on these shocks is linear in the shock: $x^*(\varepsilon, h) = \max\{0, \varepsilon - \bar{\varepsilon}(h)\}$. Thus the distribution of medical expenditures x coincides with that of the shocks themselves, above the endogenous health-specific threshold $\bar{\varepsilon}(h)$. French and Jones (2004) argue that the cross-sectional distribution of health care costs³⁷ can best be fitted by a log-normal distribution (truncated at the upper tail). We therefore assume that the health shocks ε follow a truncated log-normal distribution:

$$f(\varepsilon; \mu_\varepsilon, \sigma_\varepsilon, \underline{\varepsilon}, \bar{\varepsilon}) = \frac{\frac{1}{\varepsilon\sigma_\varepsilon} \phi\left(\frac{\ln \varepsilon - \mu_\varepsilon}{\sigma_\varepsilon}\right)}{\Phi\left(\frac{\ln \bar{\varepsilon} - \mu_\varepsilon}{\sigma_\varepsilon}\right) - \Phi\left(\frac{\ln \underline{\varepsilon} - \mu_\varepsilon}{\sigma_\varepsilon}\right)}$$

where ϕ and Φ are standard normal pdf and cdf. We then choose the mean and standard deviation $(\mu_\varepsilon, \sigma_\varepsilon)$ of the shocks such that the endogenously determined mean and standard deviation of medical expenditures in the model matches the mean and standard deviation of health expenditures for those with ε -shocks from the data.

For the catastrophic health shock z , apart from the probability of receiving it (which was determined in section 6.2.2), only the mean expenditures $\mu_z(h)$ matter. We use the percentage of labor income spent on catastrophic medical expenditures, conditional on health status h , to determine these.

Health Effort Preference Parameters and Value of Health at Terminal Date With estimates of the production function and health shock distributions in hand we now use effort data and the dynamic part of the model to calibrate parameters for exercise preference and value of health at terminal date.

We assume that the effort utility cost function takes the form

$$\gamma(h)q(e) = \gamma(h) \left[\frac{1}{1-e} - (1+e) \right]^\psi.$$

The functional form for q guarantees that $q''(e) > 0$, that $q(0) = q'(0) = 0$ and that $\lim_{e \rightarrow 1} q'(e) = \infty$, as long as $\psi > 0.5$ (which will hold true in our estimation results). The parameter ψ controls the elasticity of effort with respect to disutility, $\frac{dq}{de} \frac{e}{q}$. When ψ increases, the elasticity of the cost of effort increases for all effort levels. Because higher education raises the value of being health, we are able to use the variation in exercise levels between the high and low education groups to determine ψ . We normalize the value of γ for Fair health ($\gamma(1)$) to 1, and find relative values of γ for other health levels.

As discussed above, absent direct benefits from better health upon retirement households in the model have no incentive to exert effort, whereas in the data we still see a significant amount of exercise for those of ages 60 to 65. By introducing a terminal, education and health dependent continuation utility $v_{T+1}(educ, h)$ this problem can be rectified. Given the structure of the model and the parametric form of the health transition function $Q(h'|h, e)$ only the differences in the continuation values matter for the choice of optimal effort in the last period T . Thus, we proceed with a normalization of $v_{T+1}(< HS, h_1) = 0$.

We choose these exercise-related parameter values $\{\psi, \gamma(h)_{h=2,3,4}, v_{T+1}(educ, h)\}$ such that the model reproduces the health-contingent average effort levels of the 24-29 year olds, the 60-65 year olds, and education-contingent average effort levels. The data targets and associated model parameters are summarized in Tables

³⁶Since there is more variation in the data for earnings than for health spending by age we decided to use a finer age grouping when estimating $A(t, educ)$ using wage data than when estimating $\xi(a, educ)$ and $\phi(a, educ)$ using health (expenditure) data.

³⁷They use HRS and AHEAD data. Health care costs include health insurance premia, drug costs and costs for hospital, nursing home care, doctor visits, dental visits and outpatient care.

9 and 10. The calibrated parameter values are reported in Table 11 in Appendix F.4, together with their performance in matching the empirical calibration targets.

6.3 Model Fit

Our model is fairly richly parameterized (especially along the production function/labor earnings dimension). It is therefore not surprising that it fits health transitions, health expenditures and life cycle earnings profiles well. Appendix F.5 displays the maximum likelihood estimation results for the health transition matrix Q and shows that it fits the data well in this dimension.³⁸

In the calibration we have targeted effort levels for very young and very old households (the latter by health status), but have *not* used data on h -specific effort levels (apart from at the final pre-retirement age) in the estimation. How well the model captures the age-effort dynamics is therefore an important “test” of the model. Figures 5 (for mean effort) and 27-30 in appendix G.1 (for effort by health status) plot the evolution of effort (exercise) over the life cycle both in the data and in the model. The dotted lines show the one-standard deviation confidence bands. From Figure 5 we see that our model fits the slightly declining average exercise level over the life cycle very well, and Figures 27- 30 show the same to be true for effort conditional on health h . The only modest exception is that the model predicts somewhat of an increase in effort levels for those with excellent health towards the end of working life (to maintain excellent health in retirement) whereas the same is not true in the data. But the model-generated conditional effort life cycle plots stay easily within the one-standard deviation confidence bands (which are admittedly quite wide, on account of smaller sample size once conditioning the data both on age and health).

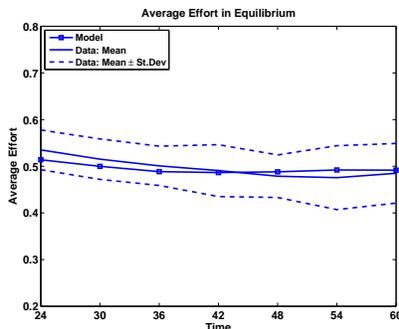


Figure 5: Average Effort in Model and Data

7 Results of the Policy Experiments: Insurance, Incentives and Welfare

After having established that the model provides a good approximation to the data for the sample period in the absence of non-discrimination policies, we now use it to answer the main counterfactual question of this paper, namely, what are the effects of introducing these policies (one at a time and in conjunction) on aggregate health, consumption and effort, their distribution, and ultimately, on social welfare.

The primary benefit of the non-discrimination policies is to provide consumption insurance against bad health, resulting in lower wages and higher insurance premia in the competitive equilibrium. However, these policies weaken incentives to exert effort to lead a healthy life, and thus worsen the long run distribution of health, aggregate productivity and thus consumption. In the next two subsections, we present the key quantitative indicators measuring this trade-off: first, the insurance benefits of policies, and second, the adverse incentive effects on aggregate production and health. Then, in subsection 7.3, we display the welfare consequences of our policy reforms. In the main text we focus on weighted averages of the aggregate

³⁸In appendix G.3 we also show that our estimation, calibration and thus welfare results are qualitatively and to large part quantitatively very robust to using various subsamples of the PSID data rather than the full data set.

variables and welfare measures across workers of different education (*educ*) types. We conclude this section with various sensitivity analyses in subsection 7.4.

7.1 Insurance Benefits of Policies

Turning first to the consumption insurance benefits of both policies, we observe from figure 6 that the combination of both policies is indeed effective in providing perfect consumption insurance. As in the first period of the constrained efficient allocation *within-group* consumption dispersion, as measured by the coefficient of variation, is zero for all periods over the life cycle if both a no-prior conditions law and a no-wage discrimination law are in place (the lines for the social planner solution and the equilibrium under both policies lie on top of one another and are identically equal to zero).³⁹ This is of course what the theoretical analysis in sections 4 and 5 predicted. Also notice from figure 6 that a wage non-discrimination law alone goes a long way towards providing effective consumption insurance, since the effect of differences in health levels on wage dispersion is significantly larger than the corresponding dispersion in health insurance premia. Thus, although a no-prior conditions law in isolation provides some consumption insurance and reduces within-group consumption dispersion by about 30%, relative to the unregulated equilibrium, the *remaining* health-induced consumption risk remains significant.

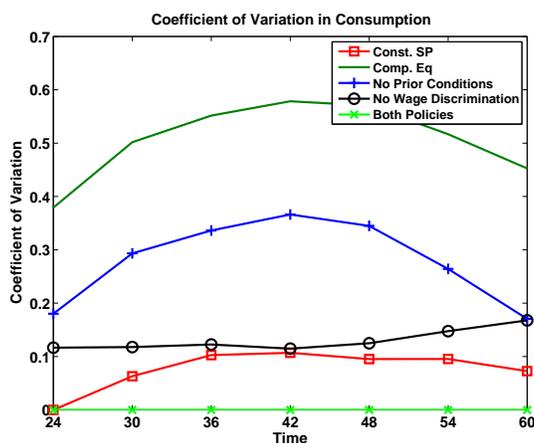


Figure 6: Consumption Dispersion

Another measure of the insurance benefits provided by the non-discrimination policies is the level of cross-subsidization or implicit transfers: workers do not necessarily pay their own competitive (actuarially fair) price of the health insurance premium or/and they are not fully compensated for their productivity. Under no-prior conditions policy, as established theoretically in Proposition 11, the healthy workers subsidize the premium of the unhealthy. Similarly, wages of the unhealthy workers are subsidized by the healthy, productive workers under the no-wage discrimination policy. Moreover, under both policies, there is cross-subsidization in both health insurance premia and wages.

Figures 7 and 8 plot the degree of cross-subsidization over the life cycle, both for households with excellent and those with fair health, and Table 14 in appendix G.2 summarizes the transfers for all health groups. The plots for the health insurance premium measures the differences between the actuarially fair health insurance premium a particular health type household would have to pay and the actual premium paid in the presence of either a no-prior conditions policy or the presence of both policies. Similarly, the wage plots display the difference between the productivity of the worker (and thus her wage in the unregulated equilibrium) and the wage received under a no-wage discrimination policy and in the presence of both policies. Negative numbers imply that the worker is paying a higher premium, or is paid lower wage than in a competitive equilibrium without government intervention. Thus such a worker, in the presence of government policies,

³⁹Due to the presence of heterogeneity in education levels the economy as a whole displays non-trivial consumption dispersion even in the presence of both policies (as it does in the solution of the restricted social planner problem).

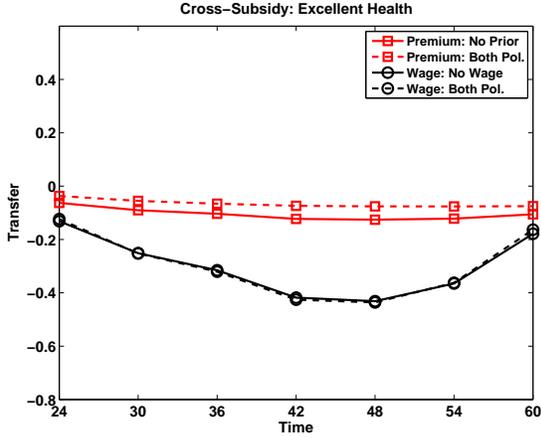


Figure 7: Cross Subsidy: Excellent Health

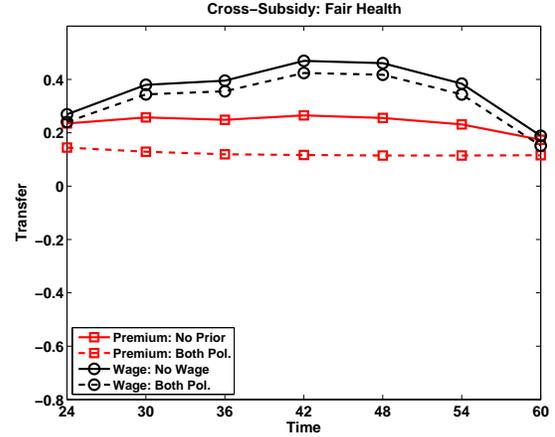


Figure 8: Cross Subsidy: Fair Health

has to transfer resources to workers of different (lower) health types. Reversely, positive numbers imply that a worker is being subsidized, i.e., she is paying a lower premium and is paid higher wage.

We observe from Figure 7 that the workers with excellent health significantly cross-subsidize the other workers, both in terms of cross-subsidies in health insurance premia as well as in terms of wage transfers. To interpret the numbers quantitatively, note that the wage transfers delivered by this group amount to about 15% of average consumption when young and close to 40% in prime working age (note that the share of workers in excellent health in the population has shrunk at that age, relative to when this cohort of workers was younger). From figure 7 we also observe that the implicit transfers induced by a no-prior conditions law are still significant (they amount to 4-8% of consumption for young workers of excellent health, and 10-15% when old), but quantitatively smaller than those implied by wage-nondiscrimination legislation.

Figure 8 displays the same plots for households of fair health. These households are the primary recipients of the transfers from workers with excellent health,⁴⁰ and for this group (which is small early in the life cycle but grows over time) the transfers are massive. In terms of their average competitive equilibrium consumption, the implicit health insurance premium subsidies amount to a large 20-30% of average consumption and the wage transfers amount to a massive 30-40% of pre-policy average consumption (and a much larger fraction of consumption of this group). These numbers indicate that the insurance benefits from both policies, and specifically from the wage nondiscrimination law, will be substantial.

An interesting property of the subsidies is that the level of subsidization implied by a given policy is higher when only one of the non-discrimination laws is enacted, relative to when both policies are present. This is especially true for the no-prior conditions law and is due to the fact that the government insures the workers with bad health through an inefficient level of medical expenditure.

Thus far, we have discussed the insurance benefits of the non-discrimination policies. In the next subsection, we analyze the aggregate dynamic effects of the policies on production and the health distribution.

7.2 Adverse Incentive Effects on Aggregate Production and Health

The associated incentive costs from each policy are inversely proportional to their consumption insurance benefits, as figure 9 shows. In this figure we plot the average exerted effort over the life cycle, in the constrained efficient and the equilibrium allocations under the various policy scenarios. In a nutshell, effort is highest in the unregulated equilibrium, positive under all policies,⁴¹ but substantially lower in the presence

⁴⁰Table 14 in the appendix shows that households with *very good* health are also called upon to deliver transfers, albeit of much smaller magnitude, and workers with *good* health are on the receiving side of (small) transfers due to wage pooling (but on the giving side of health insurance premium pooling). As the cohort ages the share of households in these different health groups shifts, and towards the end of the life cycle the now larger group of households with fair health receives subsidies from all other households, at least with respect to health insurance premia.

⁴¹Recall that, relative to the theoretical analysis, we have introduced a terminal value of health which induces not only effort in the last period even under both policies, but through the continuation values in the dynamic programming problem,

of the non-discrimination laws. More precisely, two key observations emerge from figure 9. First, the policies that provide the most significant consumption insurance benefits also lead to the most significant reductions in incentives to lead a healthy life. It is the very dispersion of consumption due to health differences, stemming from health-dependent wages and insurance premia that induce workers to provide effort in the first place, and thus the policies that reduce that consumption dispersion the most come with the sharpest reduction in incentives.⁴² Whereas a no-prior conditions law alone leads to only a modest reduction of effort, with a wage nondiscrimination law in place the amount of exercise household find optimal to carry out shrinks more significantly. Finally, if both policies are implemented simultaneously the *only* benefit from exercise is a better distribution of post-retirement continuation utility, and thus effort plummets strongly, relative to the competitive equilibrium.

The second observation we make from figure 9 is that the impact of the policies on effort is most significant at young and middle ages, whereas towards retirement effort levels under all policies converge. This is owed to the fact that the direct utility benefits from better health materialize at retirement and are independent of the nondiscrimination laws (but heavily discounted by our impatient households), whereas the productivity and health insurance premium costs from worse health accrue through the entire working life and are strongly affected by the different policies.⁴³

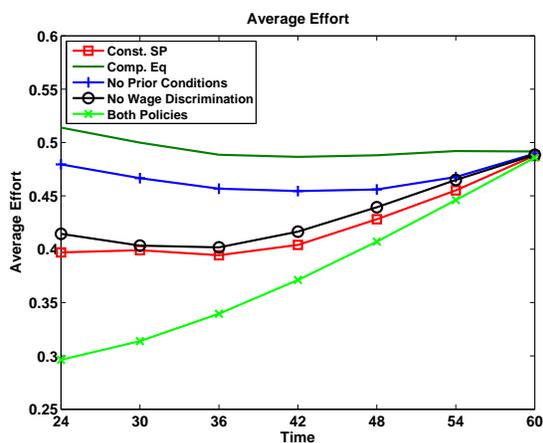


Figure 9: Effort

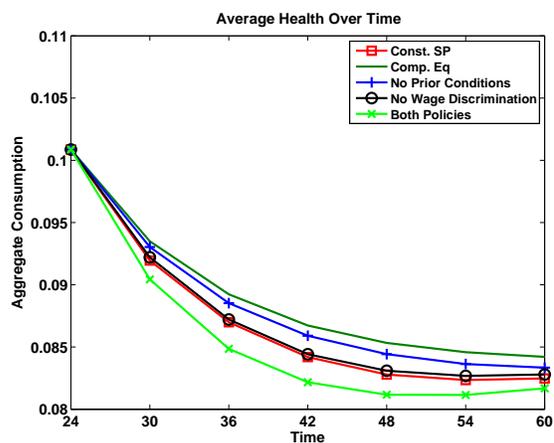


Figure 10: Average Health

Given the dynamics of effort over the life cycle (and a policy invariant initial health distribution), the evolution of the health distribution is exclusively determined by the health transition function $Q(h'; h, e)$. Figure 10 which displays average health in the economy under the various policy scenarios is then a direct consequence of the effort dynamics from Figure 9. It shows that health deteriorates under all policies as a cohort ages, but more rapidly if a no-prior conditions law and especially if a wage nondiscrimination law is in place. As with effort, the conjunction of both policies has the most severe impact on public health.

Figure 12 demonstrates that the decline of health levels over the life cycle also induce higher expenditures on health (insurance) later in life. The level of these expenditures (and thus their relative magnitudes across different policies) are determined by two factors, a) the health distribution (which evolves differently under alternative policy scenarios) and b) the equilibrium health insurance and expenditure contracts, which are fully characterized by the thresholds $\bar{\varepsilon}(h)$ from the static analysis of the model and that vary across policies. The evolution of health is summarized by figure 10, and figure 11 displays the health dependent thresholds $\bar{\varepsilon}(h)$ for the youngest households.⁴⁴ Recall from section 4 that the thresholds $\bar{\varepsilon}(h)$ under the unregulated competitive equilibrium and in the presence of both policies are socially efficient and thus the

positive effort in all periods. How quantitatively important this effect is for younger households depends significantly on the time discount factor β .

⁴²This also explains why average effort is lower in the constrained-efficient allocation relative to the equilibrium allocation.

⁴³In fact, absent the terminal (and policy invariant) direct benefits from better health the differences in effort levels across policies remain fairly constant over the life cycle.

⁴⁴The figures are qualitatively similar for older cohorts.

three graphs completely overlap. Also observe that, relative to the efficient expenditure allocation, under the no-prior conditions law workers with low health are strongly over-insured (they have *lower* thresholds, $\bar{\varepsilon}^{NP}(h_i) > \bar{\varepsilon}^{SP}(h_i)$ for $i = 1, 2$) and workers with very good and excellent health are slightly under-insured. This was the content of Proposition 11, and it is quantitatively responsible for the finding that health expenditures are highest under this policy. The reverse is true under a no-wage discrimination law: low health types are under-insured and high types are over-insured, but these differences are minor.

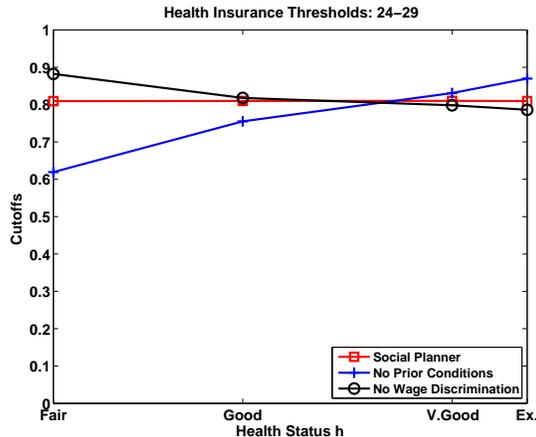


Figure 11: Cutoffs

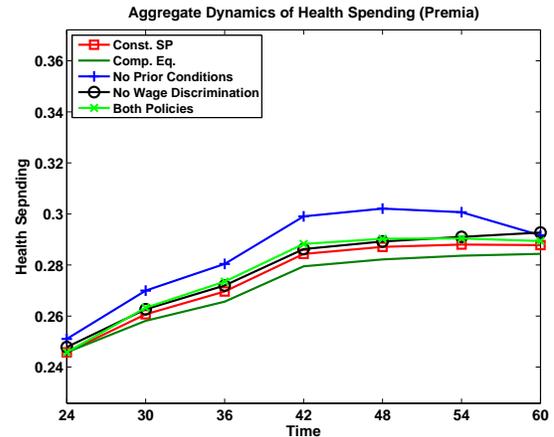


Figure 12: Health Spending

Finally, figures 13 and 14 display aggregate production and aggregate consumption over the life cycle. Since the productivity of each worker depends on her health and on the non-treated fraction of her health shock, aggregate output is lower, *ceteris paribus*, under policy configurations that lead to a worse health distribution and that leave a larger share of health shocks ε untreated. From figure 13 we observe that the deterioration of health under a policy environment that includes a wage nondiscrimination policy is especially severe, in line with the findings from figure 10. Interestingly, the more generous health insurance (for those of fair and good health) under a no-prior conditions law alone leads to output that even slightly exceeds that in the unregulated equilibrium, despite the fact that the health distribution under that policy is (moderately) worse. But health expenditures of course command resources that take away from private consumption, and as figure 14 shows, the resulting aggregate consumption over the life cycle under this policy is substantively identical to that under the wage discrimination law (and the consumption allocation is more risky under the no-prior conditions legislation). Relative to the unregulated equilibrium both policies thus entail a significant loss of average consumption in society, but for very different reasons: in the case of the wage nondiscrimination law less is produced, in the no prior conditions case more resources are spent on productivity enhancing health goods. Both effects are compounded if both policies are introduced jointly.

Overall, the effect on aggregate effort, health, production and thus consumption suggests a quantitatively important trade-off between consumption insurance and incentives. Within the spectrum of all policies, the unregulated equilibrium provides strong incentives at the expense of risky consumption, whereas a policy mix that includes both policies provides full insurance at the expense of a deterioration of the health distribution. The effects of the no-prior conditions law on both consumption insurance and incentives are modest, relative to the unregulated equilibrium. In contrast, implementing a no wage discrimination law or both policies insures away most of the consumption risk, but significantly reduces (although does not eliminate completely) the incentives to exert effort to lead a healthy life, especially early in the life cycle. In the next subsection we will now document how these two quantitatively sizable but countervailing effects translate into welfare consequences from hypothetical policy reforms.

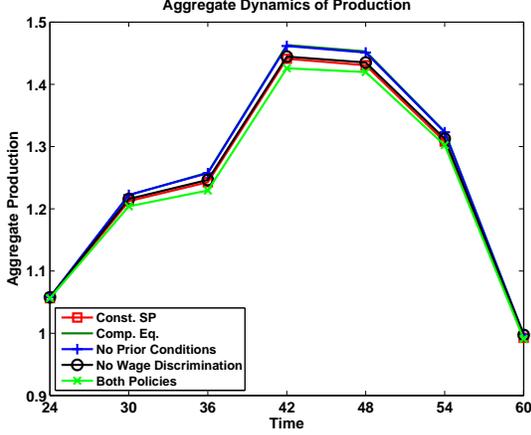


Figure 13: Production

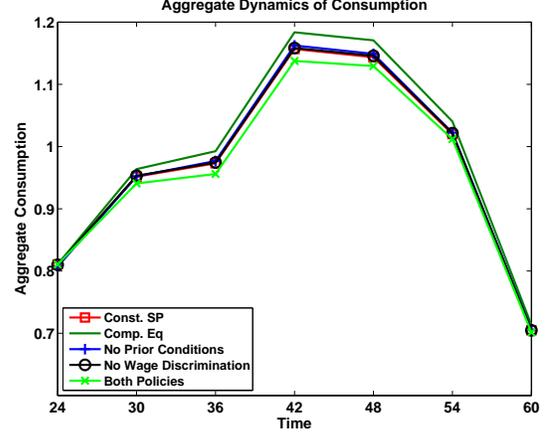


Figure 14: Consumption

7.3 Welfare Implications

7.3.1 Aggregate Welfare

In this section we quantify the welfare impact of the policy innovations studied in this paper. For a fixed initial distribution $\Phi_0(h)$ over health status,⁴⁵ denote by $W(c, e)$ the expected lifetime utility of a cohort member (where expectations are taken prior to the initial draw h of health) from an arbitrary allocation of consumption *and* effort over the life cycle.⁴⁶ Our consumption-equivalent measure of the welfare consequences of a policy reform is given by

$$W(c^{CE}(1 + CEV^i), e^{CE}) = W(c^i, e^i)$$

where $i \in \{SP, NP, NW, Both\}$ denotes the different policy scenarios. Thus CEV^i is the percentage reduction of consumption in the competitive equilibrium *consumption* allocation required to make households indifferent (ex ante) between the competitive equilibrium allocation⁴⁷ and that arising under regime i .

In order to emphasize the importance of the dynamic analysis in assessing the normative consequences of different policies we also report the welfare implications of the same policy reforms in the static version of the model in section 4, taking as given the initial distribution Φ_0 . Similar to the dynamic consequences we compute the static consumption-equivalent loss (relative to the competitive equilibrium) as

$$U(c^{CE}(1 + SCEV^i)) = U(c^i)$$

where $U(c)$ is the expected utility from the period 0 *consumption allocation*⁴⁸, under the cross-sectional distribution Φ_0 , and thus is determined by the static version of the model.⁴⁹ Therefore $SCEV^i$ provides a

⁴⁵Recall that we carry out our analysis for each (*educ*)-type separately and report averages across these types. Thus in what follows Φ_0 suppresses the (policy-independent) dependence of the initial distribution on (*educ*).

⁴⁶Using the notation from section 5 for the constrained efficient allocation and for equilibrium allocations under policy i ,

$$W(c^{SP}, e^{SP}) = V(\Phi_0)$$

$$W(c^i, e^i) = \int v_0^i(h) d\Phi_0.$$

⁴⁷Recall that even the constrained social planner problem is solved for each specific (*educ*) group separately and thus also does not permit ex-ante insurance against being part of an unfavorable (*educ*)-group. We consider this restricted social planner problem because we view the results as better comparable to the competitive equilibrium allocations.

⁴⁸In the static version of the model effort is identically equal to zero in the constrained planner problem and in the equilibrium under all policy specifications, and therefore disutility from effort is irrelevant in the static version of the model.

⁴⁹Thus, using the notation from section 4

$$U(c^i) = U^i(\Phi_0) \text{ for } i \in \{SP, NP, NW, Both\}$$

clean measure of the static gains from better consumption insurance induced by the policies against which the dynamic adverse incentive effects have to be traded off.

	Static CEV^i	Dynamic CEV^i
Constrained Social Planner	5.4026	9.2026
Competitive Equilibrium	0.0000	0.0000
No Prior Conditions Law	3.7748	6.0103
No Wage Discrimination Law	5.0157	8.9928
Both Policies	5.4026	8.2911

Table 1: Aggregate Welfare Comparisons

The static welfare consequences reported in the first column of Table 1 that isolate the consumption insurance benefits of the policies under consideration are consistent with the consumption dispersion measures displayed in Figure 6. Perfect consumption insurance, as implemented in the solution to the social planner problem and also achieved if both policies are implemented jointly, is worth about 5.4% of unregulated equilibrium consumption. Each policy in isolation delivers a substantial share of these gains, with the no wage discrimination law being more effective than the no-prior conditions law in this respect.

Turning now to the main object of interest, the dynamic welfare consequences in column 2 of Table 1 paint a different picture. Consistent with the static analysis, both policies improve on the laissez-faire equilibrium, and the welfare gains are substantial, ranging from 6% to 9% of lifetime consumption. The sources of these welfare gains are improved consumption insurance (as in the static model) and reduced effort (which bears utility costs), which outweigh the reduction in average consumption these policies entail (recall Figure 14). Furthermore, as in the static model a wage nondiscrimination law dominates a no-prior conditions law. In light of Figures 6 and 14 this does not come as a surprise: both policies imply virtually the same *aggregate* consumption dynamics, but the no-prior conditions law provides substantially less consumption insurance.

But what we really want to stress is that there are crucial differences to the static analysis. First and foremost, it is *not* optimal to introduce a no-prior conditions law once a wage non-discrimination law is already in place. The latter policy already provides effective (albeit not complete) consumption insurance, and the further reduction of incentives and the associated fall in mean consumption implied by the no prior conditions law makes a combination of both policies suboptimal. The associated welfare losses of pushing social insurance too far amount to about 0.7% of lifetime consumption.⁵⁰ Finally we see that in contrast to the static case the best policy combination (a wage nondiscrimination law alone) leads to welfare losses relative to the constrained efficient allocation, although these losses are fairly modest, in the order of 0.21% of permanent consumption.⁵¹ They emerge due to inefficiently low consumption insurance, an inefficient effort allocation and an inefficient health expenditure allocation (see again Figure 11), although the latter two effects are quantitatively modest. The last effect, however, *is* quantitatively crucial in explaining why the no-prior conditions law in isolation fares worse than the wage nondiscrimination policies (and a combination of both policies, which restores efficiency in health expenditures, recall proposition 13).

Although welfare under the optimal policy configuration (the wage nondiscrimination law) is pretty close to the constrained social optimum, one may wonder how much better a government can do that has access to more general social insurance policies such as progressive wage taxes. In appendix G.4 we therefore calculate the optimal “second best” progressive (or potentially regressive) tax code, given by a constant marginal tax rate τ and a lump-sum transfer d that balances the budget. We obtain two major findings: first, the optimal policy is very progressive, with a marginal tax rate of 82%, signaling again the large scope for social insurance of health risks prevalent in our model. Second, within the restricted set of simple tax policies considered here the government can hardly do better than implementing the wage nondiscrimination policy: the optimal tax policy leads to welfare gains, relative to that policy, of only about 0.2%, see table 18 and figure 31 in the

$$U(c^{CE}) = \int U^{CE}(h)d\Phi_0.$$

⁵⁰It should be stressed that these conclusions follow under the maintained assumption that a wage nondiscrimination law is indeed fully successful in curbing health-related wage variation, and does so completely costlessly.

⁵¹In contrast, an unconstrained planner that can perfectly control effort of households can do much better than the best policy, with a welfare gap worth 5.2% of lifetime consumption.

appendix. Thus in the absence of direct controls of health-related behavior (something that is only available to an unconstrained social planner) a simple policy that prevents wage variation by health status comes close to attaining second-best welfare, whereas in addition introducing no-prior conditions legislation pushes things to far towards social insurance, and too far away from providing dynamic incentives to lead healthy lives.

7.3.2 Heterogeneity by Health Status

The welfare consequences reported in Table 1 were measured under the veil of ignorance, before workers learn their initial health level h . They mask very substantial heterogeneity in how workers feel about these policies once their initial health status in period 0 has been revealed. Given the transfers across health types displayed in Figures 7 and 8 and the persistence of health status this is hardly surprising. Table 2 quantifies this heterogeneity by reporting dynamic consumption-equivalent variation measures, computed exactly as before, but now computed after the initial health status has been materialized.

	Fair	Good	Very Good	Excellent
Constrained Social Planner	56.8377	17.9619	3.8786	-4.7778
Competitive Equilibrium	0.0000	0.0000	0.0000	0.0000
No Prior Conditions Law	32.6768	8.6687	2.4041	-0.5140
No Wage Discrimination Law	46.3706	16.8056	4.4787	-3.4505
Both Policies	58.1009	17.5350	2.8750	-6.2285

Table 2: Welfare Comparison in the Dynamic Economy Conditional on Health

Broadly speaking, the lower a worker’s initial health status, the more she favors policies providing consumption insurance. For households of *very good* health the ranking of policies coincides with that in the second column of Table 1; households with *excellent* health instead prefer the laissez faire economy (with its implied absence of implicit transfers), whereas young households with *good* and *fair* health would support the simultaneous introduction of both policies. The differences in the preference for different policy scenarios across different h -households are quantitatively *very* large: whereas fair-health types would be willing to pay 58% of laissez faire lifetime consumption to see both policies introduced, households of excellent health would be prepared to give up 6.2% of lifetime consumption to prevent exactly this policy innovation.

Interestingly, as a cohort ages the assessment of the desirability of the policies under consideration change. In Table 3 we display the dynamic consumption-equivalent variation measures, now computed for age group 42-47.⁵² We observe that now only households with *fair* health favor both policies, whereas those middle-aged households with *very good* health now join those with *excellent* health in their opposition of any policy intervention. If we translate welfare consequences of these policies into political support, as a cohort ages the opposition against far-reaching social insurance due to both policies from those who have maintained very good health (or were lucky enough not to see it deteriorate thus far) grows stronger.

	Fair	Good	Very Good	Excellent
Constrained Social Planner	57.8839	12.7492	-5.4499	-16.3704
Competitive Equilibrium	0.0000	0.0000	0.0000	0.0000
No Prior Conditions Law	28.9282	5.0248	-1.2196	-4.4659
No Wage Discrimination Law	55.1468	13.0704	-5.3811	-15.6066
Both Policies	62.9788	11.7717	-8.8629	-20.3506

Table 3: Welfare Comparison in the Dynamic Economy Conditional on Health, Age 42-47

7.4 Robustness and Sensitivity Analysis

We conclude our discussion of the model results by documenting the robustness of our main findings to a variety of model extensions and alternative calibration strategies.

⁵²The relative ranking of policies is the same for the next younger and the next two older age groups.

7.4.1 Insurance Pooling as the Benchmark

So far we have calibrated our model so that the competitive equilibrium *without any policy* matched selected moments from the data. It could be argued that the PSID data we used (from 1997 to 2003) were generated in an environment where health insurance premia were already substantially pooled across health types, especially within companies offering group health insurance, which might bias the choice of parameters. To evaluate the potential impact of this concern we now re-calibrate the model to an alternative benchmark (which likely *overstates* the extent of pooling in the data for 1997 to 2003). To do this, we assume that insurance companies pool premia for each firm, but that firms do not discriminate with respect to hiring. Firms would not have an incentive to engage in the sort of cross-health type subsidization with respect to health expenditures that is socially efficient. Therefore, we now assume that the pattern of expenditures is chosen so as to maximize the expected net output of the workers. This implies that the firms will choose the competitive expenditure schedule, $x^{CE}(\varepsilon, h)$ while the insurance premium, P , will be given by the analog of (33) under this expenditure rule. Because of this, the wage as a function of health will still be equal to the competitive wage, $w^{CE}(h)$, but the net will now be given by $w^{CE}(h) - P$. In what follows we refer to this as the insurance-pooling equilibrium⁵³ and we use it as a benchmark to re-calibrate our model.

We find that recalibrating the model with the insurance-pooling equilibrium modestly increases the incentive costs of implementing no-prior, no-discrimination or both policies relative to the competitive equilibrium. This is because the recalibrated model needs to match the effort data with smaller gains from being healthy. As a result, the effort cost function q becomes less convex, thereby raising the incentive costs associated with greater insurance since this now leads to a larger deterioration in effort and hence in the overall health distribution over time. However, these cost increases are small and as a result the policy rankings are unchanged. We report both the new estimates for the our health effort cost function and the welfare results in tables 19 and 20 of the appendix.

7.4.2 Resource Cost and Limited Effectiveness of No Wage Discrimination

So far we modeled both policies as costless to implement and perfectly effective. Given that a wage nondiscrimination law was found to be optimal we now want to investigate how costly and/or ineffective that policy must be to change our conclusion. Assume that wages under the wage nondiscrimination law are given as

$$\tilde{w}(h) = (1 - \tau)w(h) + \tau(1 - \gamma)\bar{w}$$

where γ measures the fraction of resources lost in enforcing equal wages and τ measures the effectiveness of the wage nondiscrimination law. If $\gamma = 0$ then the policy is costless and if $\tau = 1$ it is perfectly effective in insuring wage insurance; thus $(\gamma, \tau) = (0, 1)$ corresponds to perfect wage nondiscrimination.⁵⁴

Figure 32 in appendix H.2 splits the relevant part of the parameter space into three regions and shows (not surprisingly) that as long as the wage nondiscrimination law is not too costly ($\gamma < 2\%$) and sufficiently effective ($\tau > 93.74\%$) our policy ranking remains intact. Roughly, speaking, as the wage nondiscrimination law becomes too ineffective (as long as it is not too expensive) the optimal policy configuration is to implement both policies jointly. If on the other hand the wage nondiscrimination law absorbs too much resources it is optimal to transit to a no prior condition legislation only.

7.4.3 Mismeasured Effort Inputs

We approximated our theoretical effort variable with observed exercise and (non-)smoking activities in the data. Of course these are not the only potentially health-enhancing activities households could engage in. In this subsection we briefly argue that the presence of unobserved additional effort inputs can be modeled as (non-classical) measurement error in the effort variable. Assume that true effort is given by

$$e = \lambda\tilde{e} + (1 - \lambda)e^* \tag{43}$$

⁵³This equilibrium is distinct from no-prior equilibrium because of the difference in the $x(\varepsilon, h)$ schedule.

⁵⁴For each (γ, τ) combination the health expenditure thresholds under the three policy combinations are determined as described in section 4 of the paper.

where $\lambda \in [0, 1]$ is a parameter, \tilde{e} is the object that we observe in the data and e^* are additional but unmeasured health effort inputs. We assume that the cost of producing true effort e with inputs \tilde{e}, e^* is given by $C(\tilde{e}, e^*, \chi)$ where χ is a household-specific cost parameter vector with population distribution $F(\chi)$. If $\chi = (\rho, \nu)$, and the cost function takes CES form (\tilde{e}, e^*) , then (see appendix H.3), the optimal choices of minimizing $C(\tilde{e}, e^*, \chi)$ subject to equation 43 are both proportional to e and we can rewrite 43 as

$$e = \eta\lambda\tilde{e} \tag{44}$$

where the random variable $\eta =$ has a cross-sectional distribution determined by the population distribution $F(\chi)$. We assume that that distribution is such that η is a nonnegative uniform random variable symmetrically distributed around 1. Since in our theory $e \in [0, 1]$ and in the data $\tilde{e} \in [0, 1]$ we require the upper point in the support of η to be $1/\lambda$, and by symmetry the lower support of η to be $2 - 1/\lambda \geq 0$.⁵⁵ Thus $\eta \sim UNI[2 - 1/\lambda, 1/\lambda]$ and λ measures how noisy a proxy observed effort \tilde{e} is for true effort e . If $\lambda = 1$, the distribution of η is degenerate, and observed effort is a perfect proxy for true effort. The lower is λ , the more measurement error in e we permit.

In appendix H.3 we discuss how to estimate the health transition function Q in the presence of this non-classical measurement error and plot the associated fit in figure 33. In table 4 below we show that the potential mismeasurement of effort inputs affects the magnitude of the welfare gains from the various policies somewhat, but leaves their ranking unchanged. In the first column of table 4 we recall our benchmark results, and columns 2 and 3 show results obtained under measurement error with $\lambda = 0.75$, both in case we retain the other parameter values and in case we recalibrate the model in response to the newly estimated Q .

With the re-estimated Q , health status tomorrow responds more (in expectation) to effort today, which, holding parameter values fixed, leads to larger adverse incentive effects from the policies, and thus somewhat lower welfare gains from them (compare column 2 to column 1 in table 4). The ranking of policies remains solidly intact however. When we re-calibrate the other model parameters so that under the new Q the model attains the old calibration targets, the resulting new cost function q function is less convex, thereby further increasing the adverse incentive effects from the policies and reducing the welfare benefits from them slightly (see column 3 of table 4). The main conclusion we draw from this analysis is that our key results are qualitatively and quantitatively, robust to the measurement error in health effort model here.

	$\lambda = 1$ (Benchmark)	$\lambda = 0.75$	
		Original Calibration	Re-calibrated
Constrained Social Planner	9.20	8.70	7.88
Competitive Eq	0.00	0.00	0.00
No Prior	6.01	5.66	5.32
No Wage	8.99	8.46	7.81
Both Policies	8.29	7.72	6.97

Table 4: Aggregate Welfare Comparisons: Dynamic CEVs

7.4.4 Modeling Uninsured Households

Here we briefly discuss how to augment our theory to model uninsured households. With actuarially fair insurance contracts and costs generated only by medical expenditures, all households will find it efficient to buy insurance. To rationalize the uninsured, we therefore move away from these assumptions. Suppose that production firms do not offer health insurance directly, but, as in proposition 7, instead only offer the wage contract 16. Suppose also that individual health insurance comes with a fixed cost Γ covering administrative overhead of the health insurance companies, so that the cost of purchasing insurance is now $P(h) + \Gamma$. Households opting not to purchase health insurance can still spend out off pocket expenses $x(\varepsilon, h)$ on health. Finally, suppose that individuals have fixed productivity differences, so that their output is given by $\zeta F(h, \varepsilon - x(\varepsilon, h))$ where ζ is this individual specific deterministic productivity factor. In appendix H.4 we show that uninsured households would still find it optimal to buy the competitive insurance expenditure schedule determined by the threshold rule as in section 4.1. This means that with $\zeta = 1$, $x(\varepsilon, h) = x^{CE}(\varepsilon, h)$.

⁵⁵As long as $\lambda > 0.5$ (which we will assume), this lower support is positive.

We then use this modification of the model to examine the combinations of productivity factors and fixed costs that can dissuade an individual of health h from purchasing insurance by computing the values of (ζ, Γ) that make the individual just indifferent between buying and not buying insurance in our static model. Figure 34 in appendix H.4 plots this measure of willingness to pay for insurance against ζ for different health levels h . We find that less productive (smaller ζ) individuals and those with better current health display a lower willingness to pay.^{footnote}The willingness to pay is higher for less healthy individuals since their conditional health expenditure risk is higher. This implies that for a given distribution of productivity factors in the population, and an appropriately chosen value of the fixed cost, we can rationalize why income-poor people do not buy insurance, and why, among the poor, the healthy are the most likely to be uninsured.

The impact of adding in the uninsured to our welfare analysis would be to lower the benefits of health insurance policies that force these (otherwise uninsured) individuals to buy health insurance (such as the no-prior conditions law). Health insurance subsidies are not an efficient way of increasing the welfare level of these individuals since, given the fixed cost of insurance, it is not efficient to insure them against bad health expenditure risk or health-induced productivity risk. The relative case for a wage nondiscrimination law would thus be further strengthened.

8 Conclusion

In this paper, we studied the effect of labor and health insurance market regulations on evolution of health and production, as well as welfare. We showed that both a no-wage discrimination law (an intervention in the labor market), in combination with a no-prior conditions law (an intervention in the health insurance market) provides effective consumption insurance against health shocks, holding the aggregate health distribution in society constant. However, the dynamic incentive costs and their impact on health and medical expenditures of both policies, if implemented jointly, are large. Even though both policies improve upon the laissez-faire equilibrium, implementing them *jointly* is suboptimal (relative to introducing a wage nondiscrimination in isolation). We therefore conclude that a complete policy analysis of health insurance reforms on one side and labor market (non-discrimination policy) reforms cannot be conducted separately, since their interaction might prove less favorable despite welfare gains from each policy separately.

These conclusions rest in part on our assumption that both policies can be implemented optimally at no direct overhead or enforcement costs. To us, this assumption seems potentially more problematic for the no-wage discrimination policy than the no-prior conditions policy. One can likely implement the no-prior conditions policy through the health insurance exchanges proposed by Obama Care in which a government agency links those seeking health insurance to health insurance providers and thereby overcomes, at low cost, the incentives of the health insurance companies to cherry-pick their clients. However, a similar institution (e.g. something akin to a union hall type institution), is likely to demand higher costs, given the specificity in most worker-firm matches. Given the need to allow for decentralized matching between workers and firms, the likelihood that enforcement of the no-wage policy could occur through enforcement penalties that reduce effective worker productivity opens the door to very negative outcomes as discussed in Remark 14. In addition, since the average output produced by a worker-firm pair is much larger than the expenses involved in health insurance (both in our model as well as in the data) these costs may end up being substantial.

Finally, our analysis of health insurance and incentives over the working life has ignored several potentially important avenues through which health and consumption risk affect welfare. First, the benefits of health in our model are confined to higher labor productivity, and thus we model the investment motives into health explicitly. It has abstracted from an explicit modeling of the benefits better health has on survival risk, although the positive effect of health h on the continuation utility after retirement partially captures this effect in our model, albeit in a fairly reduced form.⁵⁶ Finally, in our analysis labor income risk directly translates into consumption risk, in the absence of household private saving. We conjecture that the introduction of self-insurance via precautionary saving against this income risk further weakens the argument in favor of the policies studied in this paper. Future work has to uncover whether such an extension of the model also affects, quantitatively or even qualitatively, our conclusions about the *relative* desirability of these policies.

⁵⁶It also captures, arguably somewhat indirectly (through its stochastic impact of current health on retirement health), the utility benefits from better health during working life.

References

- [1] Acemoglu, D. and J. Angrist (2001), “Consequences of Employment Protection? The Case of the Americans with Disabilities Act.” *Journal of Political Economy*, Volume 109, No. 5, 915-957.
- [2] Aizawa, N. and H. Fang (2013), “Equilibrium Labor Market Search and Health Insurance Reform.” *NBER Working Paper No. 18698*.
- [3] Ales, L., R. Hosseini, and L. Jones (2012), “Is There “Too Much” Inequality in Health Spending Across Income Groups?” *Working Paper*.
- [4] Attanasio, O., S. Kitao and G. Violante (2011), “Financing Medicare: A General Equilibrium Analysis,” *Demography and the Economy* (chapter 9), J. Shoven (ed.), University of Chicago Press.
- [5] Battacharya, J., K. Bundorf, N. Pace and N. Sood (2009), “Does Health Insurance Make You Fat?,” *NBER Working Paper 15163*.
- [6] Brügemann, B. and I. Manovskii (2010), “Fragility: A Quantitative Analysis of the US Health Insurance System,” *Working paper*, University of Pennsylvania
- [7] Cawley, J. (2004), “The Impact of Obesity on Wages,” *Journal of Human Resources*, 39(2), 451-474
- [8] Colman, G. and D. Dave (2013), “Physical Activity and Health,” *NBER Working Paper 18858*.
- [9] Currie, J. and B. Madrian (1999), “Health, Health Insurance and the Labor Market.” In *Handbook of Labor Economics*, Volume 3C, edited by David Card and Orley Ashenfelter, Amsterdam: North Holland, 3309-3407.
- [10] DeLeire, T. (2001). “Changes in Wage Discrimination against People with Disabilities: 1984-93.” *Journal of Human Resources*, Volume 35, No. 1, 144-158.
- [11] M. Dey and C. Flinn (2005), “An Equilibrium Model of Health Insurance Provision and Wage Determination,” *Econometrica* 73, 571-627.
- [12] Edwards, R. (2008), “Health Risk and Portfolio Choice,” *Journal of Business and Economic Statistics*, 26, 472-485.
- [13] Ehrlich, I. and G. Becker (1972), “Market Insurance, Self-Insurance, and Self-Protection,” *Journal of Political Economy*, 80, 623-648.
- [14] Ehrlich, I. and H. Chuma (1990), “A Model of the Demand for Longevity and the Value of Life Extension,” *Journal of Political Economy*, 98, 761-782.
- [15] French, E. and J. Jones (2004), “On the Distribution and Dynamics of Health Care Costs,” *Journal of Applied Economics*, 19, 705-721.
- [16] Grossman, M. (1972), “On the Concept of Health Capital and the Demand for Health,” *Journal of Political Economy*, 80(2), 223-255.
- [17] Hai, R. (2012), “Causes and Consequences of Increasing Inequality in Health Insurance and Wages,” *Working Paper*.
- [18] Hall, R. and C. Jones (2007), “The Value of Life and the Rise in Health Spending,” *Quarterly Journal of Economics*, 122(1), 39-72.
- [19] Halliday, T., H. He and H. Zhang (2012), “Health Investment over the Life Cycle,” Working Paper, University of Hawaii.
- [20] Hansen, G., M. Hsu and J. Lee (2012), “Health Insurance Reform: The Impact of a Medicare Buy-In,” *NBER Working Paper 18529*.

- [21] Hugonnier, J., F. Pelgrin and P. St-Amour (2012), “Health and (Other) Asset Holdings,” *Working Paper*.
- [22] Jeske, K. and S. Kitao (2009), “U.S. Tax Policy and Health Insurance Demand: Can a Regressive Policy Improve Welfare?,” *Journal of Monetary Economics*, 56, 210-221.
- [23] Jung, J. and C. Tran (2010), “Health Care Financing over the Life Cycle, Universal Medical Vouchers and Welfare,” *Working Paper*, Towson University.
- [24] Jung, J. and C. Tran (2014), “Market Inefficiency, Insurance Mandate and Welfare: U.S. Health Care Reform 2010,” *Working Paper*, Towson University.
- [25] Kass, N., A. Medley, M. Natowicz, S. Chandros Hull, R. Faden, L. Plantinga and L. Gostin (2007), “Access to Health Insurance: Experiences and Attitudes of Those with Genetic versus Non-genetic Medical Conditions,” *American Journal of Medical Genetic*, 143A, 707–717.
- [26] Kopecky, K. and T. Koreshkova (2012), “The Joint Impact of Social Security and Medicaid on Incentives and Welfare,” *Working Paper*.
- [27] Kowalski, A. (2012), “Estimating the Tradeoff Between Risk Protection and Moral Hazard with a Nonlinear Budget Set Model of Health Insurance,” *Working Paper*.
- [28] Laun, T. (2012), “Optimal Social Insurance with Endogenous Health,” *Working Paper*.
- [29] Madrian, B. (1994), “Employment-Based Health Insurance and Job Mobility: Is There Evidence of Job-Lock?,” *Quarterly Journal of Economics*, 109, 27-54.
- [30] Miller, S. (2011), “The Effect of Insurance on Outpatient Emergency Room Visits: An Analysis of the 2006 Massachusetts Health Reform,” *Working Paper*.
- [31] Ozkan, S. (2012), “Income Differences and Health Care Expenditures over the Life Cycle,” Working Paper, Federal Reserve Board.
- [32] Pashchenko, S. and P. Porapakarm (2012), “Quantitative Analysis of Health insurance Reform: Separating Regulation from Redistribution,” *Review of Economic Dynamics*, forthcoming.
- [33] Pijoan-Mas, J. and V. Rios-Rull (2012), “Heterogeneity in Expected Longevities,” *Working Paper*.
- [34] Prados, M. (2012). “Health and Earnings Inequality,” *Working Paper*.
- [35] Sommers, A. (2007), “Access to Health Insurance, Barriers to Care, and Service Use among Adults with Disabilities,” *Inquiry*, 43, 393-405.
- [36] Weil, D. (2007), “Accounting for the Effect of Health on Economic Growth,” *Quarterly Journal of Economics*, 122, 1265-1306.
- [37] Yogo, M. (2009), “Portfolio Choice in Retirement: Health Risk and the Demand for Annuities, Housing, and Risky Assets,” NBER *Working Paper* 15307.

Appendix for Referees and Online Publication

A Proofs of Propositions

Proposition 5

Proof. Attaching Lagrange multiplier $\mu(h)$ to equation (9) and $\lambda(h)$ to equations (10) the first order conditions read as

$$u'(w(h) - P(h)) = \lambda(h) = -\mu(h) \quad (45)$$

$$\begin{aligned} \lambda(h)F_2(h, -x(0, h)) &\leq \mu(h) \\ &= \text{if } x(0, h) > 0 \end{aligned} \quad (46)$$

$$\begin{aligned} \lambda(h)F_2(h, \varepsilon - x(\varepsilon, h)) &\leq \mu(h) \\ &= \text{if } x(\varepsilon, h) > 0 \end{aligned} \quad (47)$$

Thus off corners we have

$$F_2(h, \hat{\varepsilon} - x(\hat{\varepsilon}, h)) = F_2(h, \varepsilon - x(\varepsilon, h)) = K \quad (48)$$

for some constant K . Thus off corners $\varepsilon - x(\varepsilon, h)$ is constant in ε and thus medical expenditures satisfy the cutoff rule

$$x^{CE}(\varepsilon, h) = \max[0, \varepsilon - \bar{\varepsilon}^{CE}(h)]. \quad (49)$$

Plugging (49) into (47) and evaluating it at $\varepsilon = \bar{\varepsilon}^{CE}(h)$ yields

$$\lambda(h)F_2(h, \bar{\varepsilon}^{CE}(h)) = \mu(h). \quad (50)$$

Using this result in the second part of (45) delivers the characterization of the equilibrium cutoff levels

$$F_2(h, \bar{\varepsilon}^{CE}(h)) = -1 \text{ for all } h \in H$$

which are unique, given the assumptions imposed on F . Wages, consumption and health insurance premia then trivially follow from (9) and (10). ■

Proposition 7

Proof.

Suppose that the worker was offered compensation $\mathbf{w}(h, \varepsilon - x)$ as a function of his health status h and his productivity as given in equation 16. Then, note from that the worker's insurance choice problem can be written as

$$\begin{aligned} &\max_{x(\varepsilon, h) \geq 0} \int u[\mathbf{w}(h, \varepsilon - x(\varepsilon, h)) - P(h)] f(\varepsilon) d\varepsilon \\ &\text{s.t.} \\ P(h) &= \int x(\varepsilon, h) f(\varepsilon) d\varepsilon \end{aligned}$$

From 16 we observe that wages are constant for all $x \geq \varepsilon - \bar{\varepsilon}^{CE}(h)$ and thus $x(\varepsilon, h) \leq \varepsilon - \bar{\varepsilon}^{CE}(h)$ for all ε . Since $x(\varepsilon, h)$ is restricted to be non-negative it follows that $x(\varepsilon, h) = 0$ for all $\varepsilon \leq \bar{\varepsilon}^{CE}(h)$. Conditional on $x(\tilde{\varepsilon}, h) > 0$ for a given shock $\tilde{\varepsilon} > \bar{\varepsilon}^{CE}(h)$ the first order conditions read as

$$\begin{aligned} u'[\mathbf{w}(h, \tilde{\varepsilon} - x(\tilde{\varepsilon})) - P(h)] f(\tilde{\varepsilon}) \frac{\partial \mathbf{w}(h, \tilde{\varepsilon} - x(\tilde{\varepsilon}))}{\partial x(\tilde{\varepsilon}, h)} &= f(\tilde{\varepsilon}) \lambda(h) \\ \int u'[\mathbf{w}(h, \varepsilon - x(\varepsilon, h)) - P(h)] f(\varepsilon) d\varepsilon &= \lambda(h) \end{aligned}$$

Combining, simplifying and exploiting the fact that $\frac{\partial \mathbf{w}(h, \tilde{\varepsilon} - x(\tilde{\varepsilon}))}{\partial x(\tilde{\varepsilon}, h)} = -F_2(h, \varepsilon - x(\varepsilon, h))$ for $x(\varepsilon, h) \leq \varepsilon - \bar{\varepsilon}^{CE}(h)$ yields

$$-F_2(h, \varepsilon - x(\varepsilon, h)) = \frac{\int u'[\mathbf{w}(h, \varepsilon - x(\varepsilon, h)) - P(h)] f(\varepsilon) d\varepsilon}{u'[\mathbf{w}(h, \tilde{\varepsilon} - x(\tilde{\varepsilon})) - P(h)]}. \quad (51)$$

But the health expenditure allocation

$$x(\varepsilon, h) = \max\{0, \varepsilon - \bar{\varepsilon}^{CE}(h)\}$$

yields

$$\mathbf{w}(h, \varepsilon - x(\varepsilon, h)) = w^{CE}(h)$$

for all ε , and thus (51) becomes

$$-F_2(h, \bar{\varepsilon}^{CE}(h)) = 1$$

which is satisfied by the definition of $\bar{\varepsilon}^{CE}(h)$. Next, note that the firm's profits under the pooled wage-health insurance contract is

$$F(h, \varepsilon - x) - w^{CE}(h) = \begin{cases} F(h, \varepsilon) - w^{CE}(h) & \text{if } \varepsilon < \bar{\varepsilon}^{CE}(h) \\ F(h, \bar{\varepsilon}^{CE}(h)) - w^{CE}(h) & \text{otherwise} \end{cases}. \quad (52)$$

Now we show that under the proposed separation of wage and health insurance contracts the production firm does not need to know whether the the worker has purchased health insurance. Small shocks $\varepsilon < \bar{\varepsilon}^{CE}(h)$ are not covered in any case, and for shocks $\varepsilon \geq \bar{\varepsilon}^{CE}(h)$ net profits of the firm with worker insurance is given as

$$F(h, \bar{\varepsilon}^{CE}(h)) - w^{CE}(h)$$

and without insurance

$$\begin{aligned} & F(h, \varepsilon - x) - \mathbf{w}(h, \varepsilon - x) \\ = & F(h, \varepsilon - x) - \{w^{CE}(h) - [F(h, \bar{\varepsilon}^{CE}(h)) - F(h, \varepsilon - x)]\} \\ = & F(h, \bar{\varepsilon}^{CE}(h)) - w^{CE}(h) \end{aligned}$$

Thus net profits of the firm are the same whether the worker buys insurance for health shocks $\varepsilon \geq \bar{\varepsilon}^{CE}(h)$ or not, vacating the need for the firm to verify whether the worker has purchased health insurance elsewhere or not. ■

Proposition 8

Proof. Since exercise does not carry any benefits in the static model, trivially $e^{SP} = 0$. Attaching Lagrange multiplier $\mu \geq 0$ to the resource constraint, the first order condition with respect to consumption $c(\varepsilon)$ is

$$u'(c(\varepsilon, h)) = \lambda$$

and thus $c^{SP}(\varepsilon, h) = c^{SP}$ for all $\varepsilon \in E$ and $h \in H$. Thus, not surprisingly, the social planner provides full consumption insurance to households. The optimal health expenditure allocation maximizes this consumption

$$c^{SP} = \max_{x(\varepsilon, h)} \sum_h \left\{ g(h) [F(h, -x(0, h)) - x(0, h)] + (1 - g(h)) \int f(\varepsilon) [F(h, \varepsilon - x(\varepsilon, h)) - x(\varepsilon, h)] d\varepsilon \right\} \Phi(h)$$

Denoting by $\mu(\varepsilon, h) \geq 0$ the Lagrange multiplier on the constraint $x(\varepsilon, h) \geq 0$, the first order condition with respect to $x(\varepsilon, h)$ reads as

$$-F_2(h, \varepsilon - x(\varepsilon, h)) + \mu(\varepsilon, h) = 1$$

Fix $h \in H$. By assumption 4 $F_{22}(h, y) < 0$ and thus either $x(\varepsilon, h) = 0$ or $x(\varepsilon, h) > 0$ satisfying

$$-F_2(h, \varepsilon - x(\varepsilon, h)) = 1$$

for all ε . Thus off corners $\varepsilon - x(\varepsilon, h) = \bar{\varepsilon}^{SP}(h)$ where the threshold satisfies

$$-F_2(h, \bar{\varepsilon}^{SP}(h)) = 1. \quad (53)$$

Consequently

$$x^{SP}(\varepsilon, h) = \max [0, \varepsilon - \bar{\varepsilon}^{SP}(h)].$$

The fact that $\bar{\varepsilon}^{SP}(h)$ is increasing in h , strictly so if $F_{12}(h, y) > 0$, follows directly from assumption 4 and (53). ■

Proposition 10

Proof. Let Lagrange multipliers to equations (20) and (21) be μ and $\lambda(h)$, respectively. Then, the first order conditions are:

$$\begin{aligned} \sum_h u'(w(h) - P)\Phi(h) &= \mu \\ u'(w(h) - P)\Phi(h) &= \lambda(h) \\ (1 - g(h))f(\varepsilon)[-F_2(h, \varepsilon - x(\varepsilon, h))]\lambda(h) &\leq \mu(1 - g(h))f(\varepsilon)\Phi(h) \\ &= \text{if } x(\varepsilon, h) > 0 \\ g(h)[-F_2(h, -x(0, h))]\lambda(h) &\leq \mu g(h)\Phi(h) \\ &= \text{if } x(0, h) > 0 \end{aligned}$$

Thus, off-corners we have

$$F_2(h, \varepsilon - x(\varepsilon, h)) = F_2(h, \hat{\varepsilon} - x(\hat{\varepsilon}, h)) = K$$

for some constant K and the cutoff rule is determined by

$$u'(w(h) - P)[-F_2(h, \bar{\varepsilon}^{NP}(h))] = \sum_h u'(w(h) - P)\Phi(h). \quad (54)$$

Moreover, let us take the derivative of (54) with respect to h .

$$\begin{aligned} u''(w(h) - P)\frac{\partial w(h)}{\partial h}F_2 + u'(w(h) - P)\left\{F_{12} + F_{22}\frac{\partial \bar{\varepsilon}^{NP}(h)}{\partial h}\right\} &= 0 \\ u''(w(h) - P)\frac{\partial \bar{\varepsilon}^{NP}(h)}{\partial h}\frac{\partial w(h)}{\partial \bar{\varepsilon}^{NP}(h)}F_2 + u'(w(h) - P)\left\{F_{12} + F_{22}\frac{\partial \bar{\varepsilon}^{NP}(h)}{\partial h}\right\} &= 0 \\ \Rightarrow \frac{\partial \bar{\varepsilon}^{NP}(h)}{\partial h}\left\{u''(w(h) - P)F_2\frac{\partial w(h)}{\partial \bar{\varepsilon}^{NP}(h)} + u'(w(h) - P)F_{22}\right\} &= -u'(w(h) - P)F_{12} \end{aligned}$$

Note that as $\bar{\varepsilon}$ increases $w(h)$ decreases, since $F(h, \varepsilon - x(\varepsilon, h))$ is decreasing for $\varepsilon < \bar{\varepsilon}$, and constant for $\varepsilon \geq \bar{\varepsilon}$. Thus, we have

$$\frac{\partial \bar{\varepsilon}^{NP}(h)}{\partial h} > 0.$$

■

Proposition 11

Proof. From (22), we immediately obtain

$$-F_2(h, \bar{\varepsilon}^{NP}(h)) = \frac{\sum u'(w(h) - P)\Phi(h)}{u'(w(h) - P)} \begin{array}{l} < 1 \\ = 1 \\ > 1 \end{array} \Rightarrow \begin{array}{l} \bar{\varepsilon}^{NP}(h) < \bar{\varepsilon}^{SP}(h) \\ \bar{\varepsilon}^{NP}(h) = \bar{\varepsilon}^{SP}(h) \\ \bar{\varepsilon}^{NP}(h) > \bar{\varepsilon}^{SP}(h) \end{array}$$

as $-F_2(h, \bar{\varepsilon}^{SP}(h)) = 1$.

Let us take $h_L < \tilde{h} < h_H$, and suppose

$$-F_2(h_L, \bar{\varepsilon}^{NP}(h_L)) > 1 > -F_2(h_H, \bar{\varepsilon}^{NP}(h_H)), \quad (55)$$

i.e.

$$\begin{aligned} \bar{\varepsilon}^{NP}(h_H) < \bar{\varepsilon}^{SP}(h_H) &\Rightarrow w^{NP}(h_H) > w^{SP}(h_H) \\ \bar{\varepsilon}^{NP}(h_L) > \bar{\varepsilon}^{SP}(h_L) &\Rightarrow w^{NP}(h_L) < w^{SP}(h_L), \end{aligned}$$

where $w^{SP}(h) = g(h)F(h, 0) + (1 - g(h)) \int f(\varepsilon)F(h, \varepsilon - x(\varepsilon, h))d\varepsilon$. Then, we have

$$u^{NP}(c(h_H) - P) < u^{SP}(c(h_H) - P) < u^{SP}(c(h_L) - P) < u^{NP}(c(h_L) - P),$$

where the second inequality follows from (58). This result, in combination with (55) implies

$$u'^{NP}(c(h_L) - P)[-F_2(h_L, \bar{\varepsilon}^{NP}(h_L))] > u'^{NP}(c(h_H) - P)[-F_2(h_H, \bar{\varepsilon}^{NP}(h_H))],$$

a contradiction to (22). ■

Proposition 16

Proof. Is by backward induction. Trivially $e_T(h) = 0$. In period T , since both policies are in place, the wage and health insurance premium of every household is independent of h . Thus

$$v_T(h) = u(w_T - P_T) = v_T$$

and therefore the terminal value function is independent of h . Now suppose for a given time period t the value function v_{t+1} is independent of h . Then from the first order condition with respect to $e_t(h)$ we have

$$q'(e_t(h)) = \beta v_{t+1} \sum_{h'} \frac{\partial Q(h'; h, e)}{\partial e}$$

But since for every e and every h , $Q(h'; h, e)$ is a probability measure over h' we have $\sum_{h'} \frac{\partial Q(h'; h, e)}{\partial e} = 0$ and thus $e_t(h, \gamma) = 0$ for all h , on account of our assumptions on $q'(\cdot)$. But then

$$v_t(h) = u(w_t - P_t) + \left\{ -0 + \beta v_{t+1} \sum_{h'} Q(h'; h, 0) \right\} = u(w_t - P_t) + \beta v_{t+1} = v_t$$

since $\sum_{h'} Q(h'; h, 0) = 1$ for all h . Thus v_t is independent of h . The evolution of the health distributions follows from (24), and given these health distributions wages and health insurance premia are given by (39) and (41). ■

B Further Analysis of the No-Wage Discrimination Case

B.1 Health Insurance Distortions with No-Wage Discrimination

The firm's break-even condition is

$$\sum_h \left\{ g(h)F(h, 0) + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)[F(h, \varepsilon - x^{NP}(\varepsilon, h))]d\varepsilon - w(h) \right\} \Phi(h) = 0,$$

and hence on average the production level of a worker will equal his gross wage. Taking $\varepsilon_w > 0$ and $\delta > 0$ as given, workers for whom the wage limits, $\max_{h, h'} |w(h) - w(h')| \leq \varepsilon_w$, bind will be paid either more or less than their production level depending on whether the wage discrimination bound binds from above or below. The firm will optimally choose to hire less than the population share of any health type h whose wage is above their production level, and hence some of these workers will be unemployed. Since we have assume that there is no cost to working and workers pay for their own insurance, competition over health insurance will lead these workers to increase their health insurance, $x(e, h)$, so that their productivity is within ε_w of their wage $w(h)$. In the limit as $\varepsilon_w \rightarrow 0$, this implies that

$$w(h) = g(h)F(h, 0) + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)[F(h, \varepsilon - x^{NP}(\varepsilon, h))]d\varepsilon, \quad (56)$$

holds and they are fully employed, or $w(h) - P(h) = 0$. On the flip side, there will be excess demand for workers whose expected production is more than $w(h)$, they will therefore find it optimal to either lower their insurance, and in the limit as $\varepsilon \rightarrow 0$ either (56) holds they or set $x(e, h) = 0$ if they end up at corner with respect to health insurance. Assuming that neither corner binds, this implies that the no-wage discrimination policy will be undone by adjustments in the health insurance market. This motivated our assumption that the government will choose to regulate the health insurance market to prevent this outcome as part of the no-wage discrimination policy.

For health types for which the bounds do not bind, market clearing implies that

$$w(h) = g(h)F(h, 0) + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)[F(h, \varepsilon - x^{NP}(\varepsilon, h))]d\varepsilon$$

while actuarial fairness implies that

$$P(h) = (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)x^{NP}(\varepsilon, h)d\varepsilon.$$

Hence, an efficient health insurance contract for this type will maximize $w(h) - P(h) = w^{CE}(h) - P^{CE}(h)$. Since $w^{CE}(h) - P^{CE}(h)$ is increasing in h , it follows that the wage bound binds for the lowest and highest health types.

B.2 No-Wage Discrimination with Realized Penalties in Equilibrium

Here we assume that the firm must pay a cost for having wage dispersion conditional on health type or for having the health composition of its work force differ from the population average. The wage variation penalty is assumed to take the form

$$C \sum_h [w(h) - w(0)]^2 n(h),$$

since health type 0 will have the lowest wage in equilibrium, and where C is the penalty parameter and $n(h)$ is measure of type h workers the firm hires. Note that with this penalty function the penalty will apply to all workers with health $h > 0$.⁵⁷ The penalty from having one's composition deviate from the population average is given by

$$\sum_h D \left[\frac{n(h)}{\sum n(h)} - \frac{\Phi(h)}{\sum \Phi(h)} \right]^2.$$

Since these penalties are small for small deviations, it will turn out that penalty costs will be realized in equilibrium. Since both of these penalties are real we need to subtract them from production. We will assume that here too the government will regulate the insurance market to prevent workers low health status workers raising their productivity by over-insuring themselves against health risks and high health status workers lowering their productivity by under-insuring themselves.

We begin analyzing this case by assuming that the penalties for wage discrimination C and hiring discrimination D are both finite and then we examine the equilibrium in the limit as they become large. The firm takes as given the health policy of the worker and the equilibrium wage $w(h)$ and chooses the measure of each health type to hire $n(h)$ so as to maximize

$$\begin{aligned} \max_{n(h)} \sum_h & \left[g(h) [F(h, -x(0, h)) - x(0, h)] + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon) [F(h, \varepsilon - x(\varepsilon, h)) - x(\varepsilon, h)] d\varepsilon - w(h) \right] n(h) \\ & - C \sum_h [w(h) - w^*]^2 n(h) - \sum_h \left[\frac{n(h)}{\sum n(h)} - \frac{\Phi(h)}{\sum \Phi(h)} \right]^2, \end{aligned}$$

⁵⁷If we had assumed that the form of the penalty was

$$C \int_h [w(h) - w^*]^2 \psi(h)dh,$$

where w^* is the average wage, this would mean that low productivity workers are more costly and less productive, which will discourage hiring them. Hence, with this form the low productivity workers will only be employed because of the compositional penalty, which means that the hiring penalty must bind at the margin. Hence the less than average productivity workers will be in positive net supply in equilibrium, which will complicate the analysis because some of these workers will be employed and some will not be.

where w^* is taken here to mean the lowest wage. Trivially, the firm will want to hire more than the population share of any type h for whom

$$N(h) \equiv \left[g(h) [F(h, -x(0, h)) - x(0, h)] + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon) [F(h, \varepsilon - x(\varepsilon, h)) - x(\varepsilon, h)] d\varepsilon - w(h) \right] - C [w(h) - w^*]^2$$

is positive and less than the population share if $N(h)$ is negative. Since all firms share this condition, they will all choose the same relative shares of each type of worker. Since workers are willing to work so long as $w(h) - P(h) > 0$, it follows that $w(h)$ cannot be more than w^* if $N(h)$ is not positive. To see this note that there would be excess supply of type h workers and hence the labor market would not clear. Moreover, a firm would rather hire a worker of type h at $w^* - \varepsilon$ than for w^* for ε small. Hence, if $w(h) = w^*$, then $N(h) = 0$ so long as $w^* - P(h) > 0$. Hence, for the labor market to clear for each health type, either $N(h) = 0$ for type h or $N(h) > 0$ but $w(h) - P(h) = 0$. Since the government can set $x(\varepsilon, h) = 0$ which implies that $P(h) = 0$, we assume that $w(h) - P(h) > 0$ for all health types. This implies the following proposition.

Proposition 18 *If C and D are positive but finite, and $w(h) - P(h) > 0$ for all h , then in equilibrium all households are hired, all firms are representative, and the wage $w(h)$ is equal to a worker's productivity less the cost of paying him. As C gets large, $w(h)$ converges to w^* for all $h > 0$, and the health related productivity differences are consumed by the enforcement costs.*

B.3 Realized Penalties with Both Policies

Since all that workers care about is their net wage $\tilde{w}(h)$, which is also equal to their consumption, it follows that workers are indifferent over contracts that offer combinations of a gross wage $w(h)$ and medical costs $P(h)$ for which $\tilde{w}(h) = w(h) - P(h)$ is constant. Hence, it is natural to assume that the firm takes the equilibrium *net wage* function $\tilde{w}(h)$ as given and chooses the measure of each health type to hire, $n(h)$, and its health plan, $x(\varepsilon, h)$, to solve the following problem

$$\max_{n(h), x(\varepsilon, h)} \sum_h \left[g(h) [F(h, -x(0, h)) - x(0, h)] + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon) [F(h, \varepsilon - x(\varepsilon, h)) - x(\varepsilon, h)] d\varepsilon - \tilde{w}(h) \right] n(h) - C \sum_h [\tilde{w}(h) - \tilde{w}(0)]^2 n(h) - \sum_h D \left[\frac{n(h)}{\sum n(h)} - \frac{\Phi(h)}{\sum \Phi(h)} \right]^2.$$

Proposition 19 *If C and D are positive but finite, then in equilibrium all households are hired, all firms are representative, the net wage $\tilde{w}(h)$ is equal to a worker's productivity less the cost of paying him more than $\tilde{w}(0)$, and $\tilde{w}(0) = w^{CE}(0) - P(0)$. The firm optimally sets $x(\varepsilon, h) = x^{CE}(\varepsilon, h)$. As $C \rightarrow \infty$, $\tilde{w}(h) \rightarrow \tilde{w}(0)$.*

Proof. The optimality condition for $x(h, \varepsilon)$ if $\varepsilon = 0$ is

$$F(h, -x(0, h)) - 1 \leq 0$$

and if $\varepsilon > 0$ is

$$F(h, \varepsilon - x(\varepsilon, h)) - 1 \leq 0 \text{ w. equality if } x(\varepsilon, h) > 0.$$

These are the same conditions as in the competitive equilibrium.

Next, we show that $\tilde{w}(h)$ has to be increasing in h and hence $\tilde{w}(0)$ is the lowest paid type. The wage penalty is w.r.t. to the lowest paid worker type, which we denote by w^* . Given that optimum insurance is the same as in the competitive equilibrium, it follows that the net earnings per worker is $w^{CE}(h) - P^{CE}(h) - \tilde{w}(h)$, and from before $w^{CE}(h) - P^{CE}(h)$ is increasing in h . Hence, for the firm to break even

$$\sum_h [w^{CE}(h) - P^{CE}(h) - \tilde{w}(h)] n(h) - C \sum_h [\tilde{w}(h) - w^*]^2 n(h) - \sum_h D \left[\frac{n(h)}{\sum n(h)} - \frac{\Phi(h)}{\sum \Phi(h)} \right]^2 = 0,$$

and the optimality condition for $n(h)$ is

$$[w^{CE}(h) - P^{CE}(h) - \tilde{w}(h)] - C [\tilde{w}(h) - w^*]^2 - D \left[\frac{n(h)}{\sum n(h)} - \frac{\Phi(h)}{\sum \Phi(h)} \right] \left[1 - \frac{n(h)}{\sum n(h)} \right] \frac{1}{\sum n(h)} = 0.$$

This condition implies that a firm will hire more than the population share of any type h for whom

$$\tilde{N}(h) \equiv w^{CE}(h) - P^{CE}(h) - \tilde{w}(h) - C [\tilde{w}(h) - w^*]^2 > 0,$$

and less than the population share if the reverse is true. However any health type h that are not fully employed in equilibrium would have excess members who would be happy to be hired any positive wage. Hence, either type h is paid the lowest equilibrium wage or they are fully employed. Hence, any type h for whom $w(h) > w^*$ are fully employed. Any type receiving the lowest wage must be fully employed since the firm would be willing to hire more of these workers if we lowered the bottom wage by ε . Since all workers are fully employed, it follows that all firms will choose to be representative to avoid the hiring penalty, and that $\tilde{w}(0) = w^{CE}(0) = w^*$ and $\tilde{w}(h)$ is increasing h . Finally, since the marginal penalty for a deviation in a type's net wage from the economy-wide lowest type's wage is given by

$$-C [\tilde{w}(h) - \tilde{w}(0)]^2,$$

and since this cost goes to infinity as $C \rightarrow \infty$ for any positive wage gap, it follows that as C becomes large $\tilde{w}(h) \rightarrow \tilde{w}(0)$, and all of the workers are paid as if they were the lowest health status type and all of their productivity gap is absorbed by the cost of discriminating on wages. *Q.E.D.* ■

The fact that the productivity advantage of higher health status individuals is completely absorbed by the discrimination costs means that the society as a whole gets no gain from their productivity advantage. So the health expenditures that raise their productivity above the lowest type are inefficient. In addition, expenditure on the lowest health type relaxes the wage discrimination penalty on other types. So this equilibrium outcome is not socially efficient.

C Wages in the Competitive Equilibrium

To understand the implications of proposition 5 for the behavior of equilibrium wages, note that our results imply that the equilibrium competitive wage is given by

$$\begin{aligned} w^{CE}(h) &= g(h)F(h, 0) + (1 - g(h)) \int_0^{\bar{\varepsilon}^{CE}(h)} f(\varepsilon)F(h, \varepsilon - x(\varepsilon, h))d\varepsilon \\ &\quad + (1 - g(h)) \int_{\bar{\varepsilon}^{CE}(h)}^{\bar{\varepsilon}} f(\varepsilon)F(h, \bar{\varepsilon}^{CE}(h))d\varepsilon. \end{aligned}$$

Hence

$$\begin{aligned} \frac{dw^{CE}(h)}{dh} &= g'(h) \left[\begin{aligned} &F(h, 0) - \int_0^{\bar{\varepsilon}^{CE}(h)} f(\varepsilon)F(h, \varepsilon - x(\varepsilon, h))d\varepsilon \\ &- \int_{\bar{\varepsilon}^{CE}(h)}^{\bar{\varepsilon}} f(\varepsilon)F(h, \bar{\varepsilon}^{CE}(h))d\varepsilon \end{aligned} \right] \\ &\quad + g(h)F_1(h, 0) + (1 - g(h)) \int_0^{\bar{\varepsilon}^{CE}(h)} f(\varepsilon)F_1(h, \varepsilon - x(\varepsilon, h))d\varepsilon \\ &\quad + (1 - g(h)) \int_{\bar{\varepsilon}^{CE}(h)}^{\bar{\varepsilon}} f(\varepsilon)F_1(h, \bar{\varepsilon}^{CE}(h))d\varepsilon \\ &\quad + (1 - g(h)) \int_{\bar{\varepsilon}^{CE}(h)}^{\bar{\varepsilon}} f(\varepsilon)F_2(h, \bar{\varepsilon}^{CE}(h)) \frac{d\bar{\varepsilon}^{CE}(h)}{dh} d\varepsilon, \end{aligned}$$

since net effect of the change in the integrand bounds generated by $\frac{d\bar{\varepsilon}^{CE}(h)}{dh}$ is zero. Next note that our optimality condition for $\bar{\varepsilon}^{CE}(h)$, (15), implies that

$$F_{12}(h, \bar{\varepsilon}^{CE}(h))dh + F_{22}(h, \bar{\varepsilon}^{CE}(h))d\bar{\varepsilon}^{CE}(h) = 0,$$

and hence

$$\frac{d\bar{\varepsilon}^{CE}(h)}{dh} = \frac{-F_{12}(h, \bar{\varepsilon}^{CE}(h))}{F_{22}(h, \bar{\varepsilon}^{CE}(h))}.$$

This result, along with (15), implies that

$$\begin{aligned} \frac{dw^{CE}(h)}{dh} &= g'(h) \left[\begin{aligned} &F(h, 0) - \int_0^{\bar{\varepsilon}^{CE}(h)} f(\varepsilon)F(h, \varepsilon - x(\varepsilon, h))d\varepsilon \\ &- \int_{\bar{\varepsilon}^{CE}(h)}^{\bar{\varepsilon}} f(\varepsilon)F(h, \bar{\varepsilon}^{CE}(h))d\varepsilon \end{aligned} \right] \\ &+ g(h)F_1(h, 0) + (1 - g(h)) \int_0^{\bar{\varepsilon}^{CE}(h)} f(\varepsilon)F_1(h, \varepsilon - x(\varepsilon, h))d\varepsilon \\ &+ (1 - g(h)) \int_{\bar{\varepsilon}^{CE}(h)}^{\bar{\varepsilon}} f(\varepsilon)F_1(h, \bar{\varepsilon}^{CE}(h))d\varepsilon \\ &- (1 - g(h)) \int_{\bar{\varepsilon}^{CE}(h)}^{\bar{\varepsilon}} f(\varepsilon)F_2(h, \bar{\varepsilon}^{CE}(h)) \frac{F_{12}(h, \bar{\varepsilon}^{CE}(h))}{F_{22}(h, \bar{\varepsilon}^{CE}(h))} d\varepsilon. \end{aligned} \quad (57)$$

All of the terms in (57) are trivially positive except the last, which is negative since $F_{22} < 0$. However, so long as the spillover ratio F_{12}/F_{22} evaluated at $(h, \bar{\varepsilon}^{CE}(h))$ is not too negative then, then wages will vary positive with health status. Note that this is trivially implied if the direct effect of the change in health status offsets the spillover, or

$$F_1(h, \bar{\varepsilon}^{CE}(h)) - F_2(h, \bar{\varepsilon}^{CE}(h)) \frac{F_{12}(h, \bar{\varepsilon}^{CE}(h))}{F_{22}(h, \bar{\varepsilon}^{CE}(h))} > 0. \quad (58)$$

Note that this is a condition purely on the fundamentals of the economy since $\bar{\varepsilon}^{CE}(h)$ is given by an (implicit) equation that depends only on exogenous model elements. We summarize our results in the following proposition:

Proposition 20 *The competitive wage is increasing in h if (57) is positive.*

D Computation of the Social Planner Problem

The social planner problem in section can be solved numerically, either by making the problem recursive or by brute force optimization over the finite-dimensional vectors $\{c_t(h), e_t(h), V_t(h)\}$. The recursive problem of the planner has as state variable the cross-sectional distribution over health status Φ , which makes it rather cumbersome to solve. Instead, we solve the sequence problem directly, using a penalty function approach to assure that the aggregate resource constraint is satisfied in every period t . Thus the problem we solve is

$$\begin{aligned} &\max_{\{c_t(h), e_t(h), V_t(h)\}_{t=0}^T} \sum_h \Phi_0(h) V_0(h) - \sum_{t=1}^T P \left(Y(\Phi_t) - \sum_h c_t(h) \Phi_t(h) \right) \\ &s.t. \\ V_t(h) &= u(c_t(h)) - q(e_t(h)) + \beta \sum_{h'} Q(h'; h, e_t(h)) v_{t+1}(h') \end{aligned} \quad (59)$$

$$q'(e_t(h)) = \beta \sum_{h'} \frac{\partial Q(h'; h, e_t(h))}{\partial e_t(h)} v_{t+1}(h'), \quad (60)$$

$$\Phi_{t+1}(h') = \sum_h Q(h'; h, e_t(h)) \Phi_t(h). \quad (61)$$

where the penalty function P is given by

$$P(x) = \frac{\kappa}{2} (\min\{0, x\})^2 = \begin{cases} 0 & \text{if } x \geq 0 \\ \frac{\kappa}{2} (x)^2 & \text{if } x < 0 \end{cases}$$

where κ is a penalty parameter. Ideally we want κ to be large (to make sure the constraints are satisfied at the optimal solution), but the larger is κ the harder might the optimization problem be solved. This suggests the following algorithm

Algorithm 21 Choose κ^0 small. Solve the above maximization problem with κ^0 , and denote the solution as $\{c_t^0(h), e_t^0(h), V_t^0(h)\}_{t=1}^T$ and denote the solution in iteration step n as $\{c_t^n(h), e_t^n(h), V_t^n(h)\}_{t=0}^T$. Then iterate on

1. For given κ^n solve the maximization problem using $\{c_t^{n-1}(h), e_t^{n-1}(h), V_t^{n-1}(h)\}_{t=0}^T$ as initial guess. Optimal solution is $\{c_t^n(h), e_t^n(h), V_t^n(h)\}_{t=0}^T$

2. Update

$$\kappa^{n+1} = \phi \kappa^n$$

where ϕ is a computational parameter that trades off speed (high ϕ) and stability (low ϕ).

3. Stop if

$$\|c^{n+1} - c^n\| < \varepsilon.$$

E Computation of the Equilibrium with a No-Prior-Conditions Law and/or a No-Wage Discrimination Law

The algorithm to solve this version of the model shares its basic features with that for the social planner problem, but differs in terms of the sequence of variables on which we iterate:

Algorithm 22 1. Guess a sequence⁵⁸ $\{Eu'_t, P_t\}_{t=0}^T$.

2. Given the guess use equations (30)-(33) to determine health cutoffs and wages $\{\bar{\varepsilon}_t^{NP}(h), w_t(h)\}$.

3. Given $\{w_t(h), P_t\}$, solve the household dynamic programming problem (34) for a sequence of optimal effort policies $\{e_t(h)\}_{t=0}^T$.

4. From the initial health distribution Φ_0 use the effort functions $\{e_t(h)\}_{t=0}^T$ to derive the sequence of health distributions $\{\Phi_t\}_{t=0}^T$ from equation (24).

5. Obtain a new sequence $\{Eu_t'^{new}, P_t^{new}\}_{t=0}^T$ from (32) and (33).

6. If $\{Eu_t'^{new}, P_t^{new}\}_{t=0}^T = \{Eu'_t, P_t\}_{t=0}^T$ we are done. If not, go to step 1. with new guess $\{Eu_t'^{new}, P_t^{new}\}_{t=0}^T$.

The algorithm for no-wage discrimination is a slight modification of that for no-prior conditions. The algorithm iterates over $\{Eu'_t, w_t\}_{t=0}^T$. In Step 1 given the guess use equations (36)-(40) to determine health cutoffs and premia $\{\bar{\varepsilon}_t^{NP}(h), P_t(h)\}$. In Step 4 obtain a new sequence $\{Eu_t'^{new}, w_t^{new}\}_{t=0}^T$ from (39) and (38). With both policies, equation (41) replaces (40) in all expressions.

F Details for Data and Calibration

F.1 Details of the Augmented Model Analysis: Inclusion of the z -shock

We assume that households *must* incur the cost z , when the z -shock hits. This assumption and the fact that households are risk averse imply that the z -shock will be fully insured in the competitive equilibrium under any policy (and of course by the social planner).

⁵⁸Instead of $\{Eu_t\}$ one could iterate on $\{w_t(h)\}$ which is more transparent, but significantly increases the dimensionality of the problem.

Moreover, we assume that households receiving a z -shock can still work, but that their productivity is only ρ times that of a healthy worker. Therefore, in a competitive equilibrium, the wage of a worker with health status h is given by

$$w(h) = g(h)F(h, 0) + \rho\kappa(h)F(h, 0) + (1 - g(h) - \kappa(h)) \int_0^{\bar{\varepsilon}} F(h, \varepsilon - x(\varepsilon, h))f(\varepsilon)d\varepsilon$$

and the health insurance premium is determined as

$$P(h) = (1 - g(h) - \kappa(h)) \int_0^{\bar{\varepsilon}} x(\varepsilon, h)f(\varepsilon)d\varepsilon + \mu_z(h)$$

Given our assumptions there is no interaction between the z -shocks and the health insurance contract problem associated with the ε -shock since it is prohibitively costly by assumption not to bear the z -expenditures. The role of the z -expenditures is to soak up the most extreme health expenditures observed in the data associated with catastrophic illnesses, but to otherwise leave our theory from the previous sections unaffected.

The static analysis goes through completely unchanged in the presence of the z -shocks. In the dynamic analysis the benefits of higher effort e and thus a better health distribution $\Phi_t(h)$ now also include a lower probability $\kappa(h)$ of receiving a positive z -shock and a lower mean expenditure $\mu_z(h)$ from that shock with better health h . This extension of the model leads to straightforward extensions of the expressions derived in the analysis of the dynamic model in section 5, and does not change any of the theoretical properties derived in sections 4 and 5.

F.2 Descriptive Statistics of the PSID Data

Before we proceed to descriptive statistics of the PSID data, we summarize, in Table 5, the mapping between variables in our model and data.

Table 5: Mapping between Data and Model

Model	Description	Data	
		PSID Variable	Actual Data Used
x, μ_z	Medical Expenditure	Average of total expenditure reported in 1999, 2001, 2003	1997-2002
w	Earning	Average of total labor income reported in 1999, 2001, 2003	1998,2000,2002
h	Health Status	Self-reported Health in 1997	1997

Since our model period is six years, we take average of reported medical expenditure and wages over six year periods that we observe. Moreover, we use health status data from 1997 (rather than 1999) to capture the effect of health on wages and medical expenditure.

Table 6 documents descriptive statistics of key variables from the 1999 PSID data that we use in our analysis.

In the PSID, each individual (head of household) self-reports his health status in a 1 to 5 scale, where 1 is Excellent, 2, Very Good, 3, Good, 4, Fair, and 5 is Poor. Even with large number of observations, only about 1% of total individuals report their health status to be poor. Thus, for our analysis, we will use four levels of health status (merge poor and fair together).⁵⁹ Since PSID reports household medical expenditure, we control for family size using modified OECD equivalence scale.⁶⁰

As we model working-age population, each household starts his life as a 24 year old and makes economic decisions until he is 65 years old. Our model time period is 6 years and thus they live for 7 time periods. We choose six year time period to capture the effect of exercises on health transition. Since exercises tend to have positive longer-term effects than do medical expenditure, by allowing for a medium-term time period, we are able to quantify the impact of exercises in a more reliable way.

⁵⁹Labor income and medical expenditure data for fair health in Table 6 include poor (5) in data.

⁶⁰Each additional adult gets the weight of 0.5, and each child, 0.3.

Table 6: Descriptive Statistics of Key Variables in PSID

	Mean	Std. Dev.	Min	Max
Age	41	10	23	65
Labor Income	30,170	40,573	0	1,153,588
if Labor Income > 0	32,076	41,097	0.55	1,153,588
Excellent	38,755	55,406	0	940,804
Very Good	32,768	40,351	0	1,153,588
Good	25,516	25,908	0	384,783
Fair	12,605	13,926	0	81,300
Medical Expenditure	1,513	4,624	0	127,815
Excellent	1,234	2,374	0	28,983
Very Good	1,647	5,812	0	127,815
Good	1,486	4,283	0	93,298
Fair	1,792	4,950	0	65,665
Health Status	2.77	0.95	1	4
Physical Activity: fraction(number) of days in a year				
Light	0.63 (230.99)	0.39 (142.28)	0	1 (365)
Heavy	0.29 (105.69)	0.35 (126.85)	0	1 (365)
Smoking: number of cigarettes in a day	3.73	7.93	0	87

Data on Health Transitions Table 7 presents the transition matrix of health status over six years. We see that health status is quite persistent.

Table 7: Health Transition over 6 years

	Excellent	Very Good	Good	Fair	Total
Excellent	1,286 49.14 %	904 34.54 %	335 12.80 %	92 3.52 %	2,617 100 %
Very Good	482 12.63 %	1,844 48.31 %	1,217 31.88 %	274 7.18 %	3,817 100 %
Good	187 5.98 %	712 22.76 %	1,592 50.90 %	637 20.36 %	3,128 100 %
Fair	36 2.47%	109 7.47 %	358 24.52 %	957 65.55 %	1,460 100 %
Total	1,991 18.06 %	3,569 32.38 %	3,502 31.77%	1,960 17.78 %	11,022 100 %

Physical Activity Data Figures 15 – 18 plot density of light physical activity, heavy physical activity, smoking behavior, and smoking behavior conditional on smoking. Smoking is reported as number of cigarettes smoked in a day as a fraction of 100 cigarettes (maximum possible answer in PSID).

F.3 Health Shocks, Distribution of Medical Expenditures, and Discussion of Categorization of Health Shocks

Before going into discussing the medical expenditure distribution in data, we briefly discuss the appropriate counterparts of data moments for our model. In our model, households do not consume medical care when they do not get a health shock (although, they can choose not to spend any in case of health shock, since $x^*(h, \varepsilon) = \max\{0, \bar{\varepsilon}(h)\}$). Therefore, in data, we are interested in the distribution of medical expenditure conditional on having gotten any health shocks (which we have some information in PSID).

Table 8 summarizes medical expenditure by shock. Note that all numbers reported are yearly average taken over six years (1997-2002).

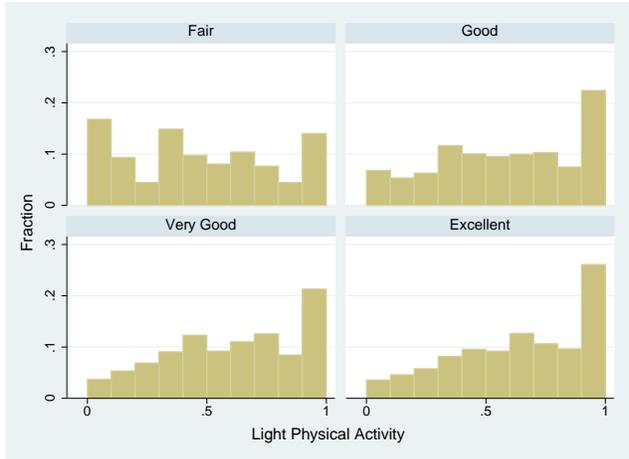


Figure 15: Density of Light Physical Activity

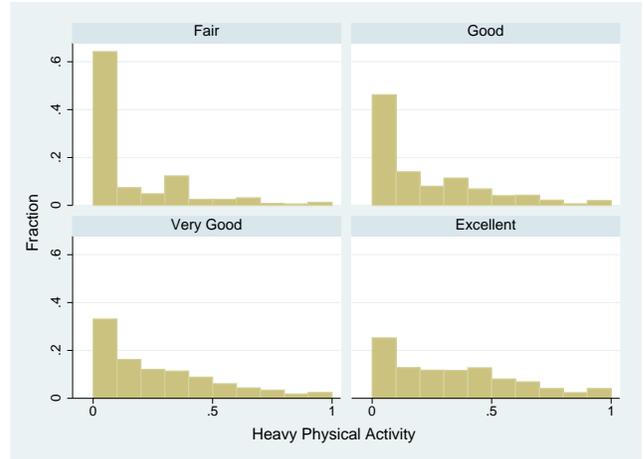


Figure 16: Density of Heavy Physical Activity

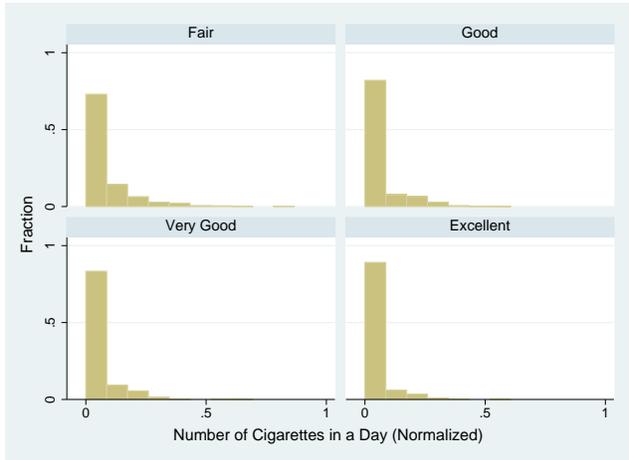


Figure 17: Density of Smoking

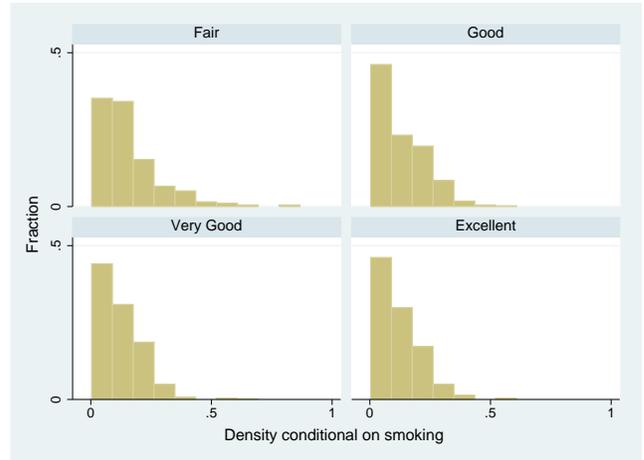


Figure 18: Conditional Density of Smoking

We see that cancer, heart attack, and heart disease incur the most medical expenditure, and thus we categorize them to be *catastrophic* shocks (z -shocks). Although the diseases PSID specifically reports information on are those that are common, they are not, by all means, exhaustive of the kind of health diseases that one can be diagnosed with. And this is hinted when we look at the medical expenditure statistics for those who report to have missed work due to illness. The maximum amount of medical expenditure they spend exceeds those of the others, and this might be due to some severe diseases for which they had to be treated.

Therefore, in addition to cancer, heart attack, and heart disease, we categorize those who have spent more than their labor income on medical expenditure as having had a *catastrophic* (z) *health shock*.⁶¹ Those who had a health shock that were not cancer, heart attack, or heart disease, and who spent less than their income on medical expenditure is considered to have had an ε -shock.⁶²

Figures 19 - 22, plot logs of medical expenditure distribution for all population, for those with ANY health shock, those with z -shock, and those with ε -shock. By definition, mean medical expenditure of z -shock households are higher than those of ε -shock, and so are standard deviations.

⁶¹Categorizing catastrophic health shocks using expenditures as percentage of income is not new. There has been discussion on insuring catastrophic health shocks, and they mostly refer to *high* amount of expenditure as percentage of income.

⁶²In PSID sample, median of percentage of labor income spent on medical expenditure is 2%, and the mean, 13%. Only about 5% of households with health shocks spend medical expenditure in excess of their labor income.

Table 8: Average Medical Expenditure by Health Shock Categories

	Obs	Mean	Std. Dev.	Min	Max
All	4,226	1,513	4,624	0	127,815
No Shock	1,419	1,350	4,447	0	101,952
Any Shock	2,807	1,595	4,710	0	127,815
Catastrophic Disease Shock	168	3,745	9,363	0	93,298
Cancer	51	5,210	15,134	0	93,298
Heart Attack	46	3,334	4,705	0	27,161
Heart Disease	94	3,382	5,535	0	38,500
Light Shock	2,767	1,585	4,732	0	127,815
Diabetes	183	2,088	7,196	0	93,298
Stroke	33	2,200	4,905	0	27,161
Arthritis	322	1,684	3,166	0	38,500
Hypertension	566	1,825	6,143	0	93,298
Lung Disease	63	1,705	2,476	0	12,595
Asthma	61	1,135	1,444	0	7,170
Ill	2,351	1,637	5,040	0	127,815
z -shock	297	4,704	12,834	0	127,815
ε -shock	2,510	1,227	2,023	0	32,909

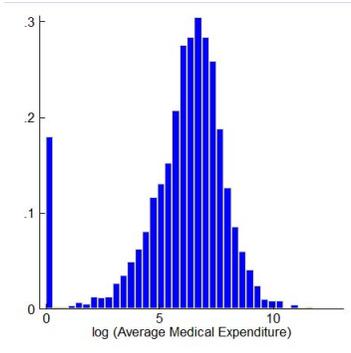


Figure 19: Average Medical Expenditure Distribution

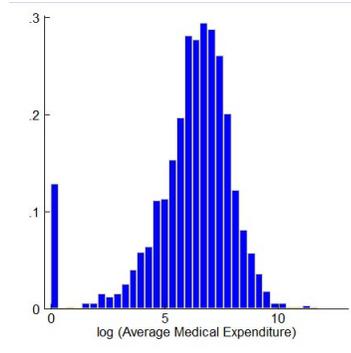


Figure 20: Average Expenditure with Health Shock

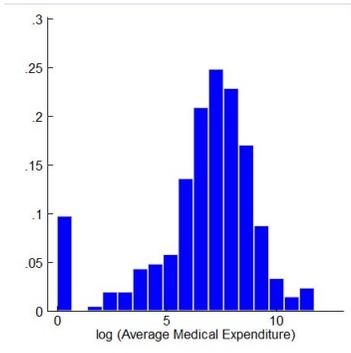


Figure 21: Average Expenditure w/ z -shock

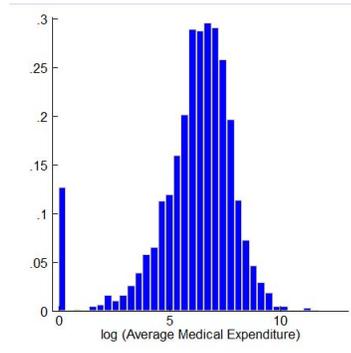


Figure 22: Average Expenditure w/ ε -shock

F.4 Calibration Results

We summarize our calibration results here. Tables 9 and 10 report data targets for each parameters to be calibrated, and table 11 documents the calibration results.

Table 9: Data Targets

	Parameters	Data Targets
Health Status	$\{h_i\}_{i=1,2,3,4}$	Income of h_i relative to h_1 $\log \frac{w(h_2)}{w(h_1)} = 0.2739$ $\log \frac{w(h_3)}{w(h_1)} = 0.4691$ $\log \frac{w(h_4)}{w(h_1)} = 0.5948$ Income of <i>Old</i> relative to <i>Young</i> $\log \frac{w(O)}{w(Y)} = 0.1114$
Production Function	$A(t, educ)$	Income in t of less than HS relative to Income of <i>Young</i> and Fair health $t = 1, < HS : -0.0042$ $t = 2, < HS : 0.1449$ $t = 3, < HS : 0.1715$ $t = 4, < HS : 0.1980$ $t = 5, < HS : 0.0907$ $t = 6, < HS : -0.0969$ $t = 7, < HS : -0.1112$ Income in t of HS Grad relative to Income of <i>Young</i> and Fair health $t = 1, HS : 0.2980$ $t = 2, HS : 0.4738$ $t = 3, HS : 0.5082$ $t = 4, HS : 0.5988$ $t = 5, HS : 0.6060$ $t = 6, HS : 0.5395$ $t = 7, HS : 0.2406$
	$\phi(a, educ)$	% Income spent on Med Exp. by Health $\frac{\mathbb{E}(x h_1)}{\mathbb{E}(w h_1)} = 0.0525$ $\frac{\mathbb{E}(x h_2)}{\mathbb{E}(w h_2)} = 0.0429$ $\frac{\mathbb{E}(x h_3)}{\mathbb{E}(w h_3)} = 0.0353$ $\frac{\mathbb{E}(x h_4)}{\mathbb{E}(w h_4)} = 0.0308$
	$\xi(a, educ)$	% Income on Med Exp. by Education and Age ($a \in \{Y, O\}$) $\frac{\mathbb{E}(x Y, < HS)}{\mathbb{E}(w Y, < HS)} = 0.0386$ $\frac{\mathbb{E}(x Y, HS)}{\mathbb{E}(w Y, HS)} = 0.0348$ $\frac{\mathbb{E}(x O, < HS)}{\mathbb{E}(w O, < HS)} = 0.0428$ $\frac{\mathbb{E}(x O, HS)}{\mathbb{E}(w O, HS)} = 0.0356$

Table 10: Data Targets (continued)

	Parameters	Data Targets
ε -shock Distribution	μ, σ_ε	Mean and St.Dev of Agg. Med. Exp. on Light Shocks $\frac{\mathbb{E}(x)}{\mathbb{E}(w)} = 0.0362$ $\frac{\sigma(x)}{\mathbb{E}(x)} = 1.6462$
z -shock Distribution	$\mu_z(h)$	% of Income Spent on Catastrophic Shock by Health $\frac{\mathbb{E}(z h_1)}{\mathbb{E}(w h_1)} = 0.4664$ $\frac{\mathbb{E}(z h_2)}{\mathbb{E}(w h_2)} = 0.2234$ $\frac{\mathbb{E}(z h_3)}{\mathbb{E}(w h_3)} = 0.1520$ $\frac{\mathbb{E}(z h_4)}{\mathbb{E}(w h_4)} = 0.1261$
Exercise Disutility and Terminal Value	$\psi, \{\gamma(h)\}_{h=2,3,4}$ $\{v_{T+1}(educ, h)\}$	Mean Exercise in $t = 1$ by Health $\mathbb{E}(e_{t=1} h_1) = 0.4648$ $\mathbb{E}(e_{t=1} h_2) = 0.4983$ $\mathbb{E}(e_{t=1} h_3) = 0.5268$ $\mathbb{E}(e_{t=1} h_4) = 0.5788$ Exercise in the Last Period by Health $\mathbb{E}(e_{t=T} h_1) = 0.3850$ $\mathbb{E}(e_{t=T} h_2) = 0.4926$ $\mathbb{E}(e_{t=T} h_3) = 0.5315$ $\mathbb{E}(e_{t=T} h_4) = 0.5400$ Exercise in the Last Period by Education $\mathbb{E}(e_{t=T} < HS) = 0.4502$ $\mathbb{E}(e_{t=T} HS) = 0.5021$ Measure of Fair Health in the Last Period $\Phi_T(h_1) = 0.1944$

Table 11: Calibrated Parameters

Parameters	Description	Value	Statistics	Data	Model
h_1	Health Status	0.0326	Relative log Wages in Health and Age	0.2739	0.2754
h_2		0.0713		0.4691	0.4732
h_3		0.1140		0.5948	0.5659
h_4		0.1294		0.1114	0.1051
$A(t = 1, < HS)$	Effect of Age, Education on Productivity	1.0000	Relative log Wages in Time and Education	-0.0042	-0.0042
$A(t = 2, < HS)$		2.6292		0.1449	0.1442
$A(t = 3, < HS)$		3.0200		0.1715	0.1705
$A(t = 4, < HS)$		3.5302		0.1980	0.1988
$A(t = 5, < HS)$		2.2891		0.0907	0.0909
$A(t = 6, < HS)$		0.3803		-0.0969	-0.0966
$A(t = 7, < HS)$		0.2555		-0.1112	-0.1106
$A(t = 1, HS)$		3.7158		0.2980	0.2926
$A(t = 2, HS)$		6.0301		0.4738	0.4376
$A(t = 3, HS)$		6.8649		0.5082	0.4675
$A(t = 4, HS)$		9.2934		0.5988	0.6417
$A(t = 5, HS)$		9.6724		0.6060	0.6515
$A(t = 6, HS)$		8.2868		0.5395	0.5726
$A(t = 7, HS)$		3.2369		0.2406	0.2454
$\phi(Y, < HS)$	Effect of Med. Exp. on Productivity	0.2826	% Income on Med.Exp. by Health,Education,Age	0.0525	0.0511
$\phi(O, < HS)$		0.3161		0.0429	0.0424
$\phi(Y, HS)$		0.2958		0.0353	0.0341
$\phi(O, HS)$		0.3180		0.0308	0.0275
$\xi(Y, < HS)$		0.0080		0.0386	0.0383
$\xi(O, < HS)$		0.0129		0.0348	0.0344
$\xi(Y, HS)$		0.0066		0.0428	0.0435
$\xi(O, HS)$		0.0444		0.0356	0.0380
μ_ε	Mean of health shock	0.9816	Mean Medical Expenditure	0.0362	0.0390
σ_ε	St. Dev. of health shock	0.0385	St.Dev Medical Expenditure	1.6462	1.6493
$\mu_z(h_1)$	Mean of z -shock	0.3548	Income spent on Catastrophic Shock	0.4664	0.4635
$\mu_z(h_2)$		0.2256		0.2234	0.2238
$\mu_z(h_3)$		0.1869		0.1520	0.1521
$\mu_z(h_4)$		0.1697		0.1261	0.1258
ψ	Exercise Disutility; Value of health at retirement	3.1454	Mean of Exercise, $t = 1$ by Health; Conditional Mean Effort in $t = T$	0.4648	0.4057
$\gamma(h_2)$		0.2086		0.4983	0.4862
$\gamma(h_3)$		0.0592		0.5268	0.5169
$\gamma(h_4)$		0.0182		0.5788	0.5655
$v_{T+1}(< HS, h_2)$		0.0096		0.3850	0.3766
$v_{T+1}(< HS, h_3)$		0.5028		0.4926	0.4684
$v_{T+1}(< HS, h_4)$		1.4985		0.5315	0.5361
$v_{T+1}(HS, h_1)$		0.2424		0.5400	0.6261
$v_{T+1}(HS, h_2)$		1.3846		0.4502	0.4459
$v_{T+1}(HS, h_3)$		2.1622		0.5021	0.5042
$v_{T+1}(HS, h_4)$		4.1319		0.1944	0.2142

F.5 Estimation Results

Health Transition Using the functional form described in the main body of the paper, we estimate the health transition function in the following way. Let set of parameters to be estimated be

$$\theta = \left\{ \{G(h, h')\}, \{\delta_l, \delta_h, \delta_s\}, \phi(h), \lambda(h), \alpha_1(h), \alpha_2(h) \right\}.$$

We use Maximum Likelihood Estimation to estimate these parameters⁶³.

We maximize the likelihood, which is given by

$$\begin{aligned} \mathcal{L}(\theta) &= \prod_{obs=1}^{N_{obs}} [Q(h'|h_{obs}, e_{obs}^l, e_{obs}^h, 1 - s_{obs})]^{1\{h'=h'_{obs}\}} \\ \log \mathcal{L}(\theta) \equiv \ell(\theta) &= \sum_{obs=1}^{N_{obs}} 1\{h' = h'_{obs}\} \log [Q(h'|h_{obs}, e_{obs}^l, e_{obs}^h, 1 - s_{obs})]. \end{aligned}$$

The estimation yields us the following estimated parameter values ($h = 1, 2, 3, 4$ corresponds to health being fair, good, very good, and excellent, respectively, i.e. the higher the h the better one's health status.).

$$\begin{aligned} \hat{G}(h, h') &= \begin{bmatrix} 0.8686 & 0.0913 & 0.0300 & 0.0101 \\ 0.6777 & 0.2380 & 0.0575 & 0.0268 \\ 0.1034 & 0.4822 & 0.2928 & 0.1216 \\ 0.0649 & 0.2523 & 0.6319 & 0.0509 \end{bmatrix} \\ \delta &= [0.3811, 0.4600, 0.1589] \\ \phi &= [2.9359, 1.0742, 0.8777, 10.2570] \\ \lambda &= [0.3947, 0.0196, 0.4601, 0.3544] \\ \alpha_1 &= [0.9788, 21.7213, 4.8434] \\ \alpha_2 &= [0.8295, 8.8962] \end{aligned}$$

The estimated transition functions are plotted in Figures 23 - 26. In the figures, the smoothed functions are estimated transition, whereas the straight lines represent the data. We see that our functional form fits the data quite well.

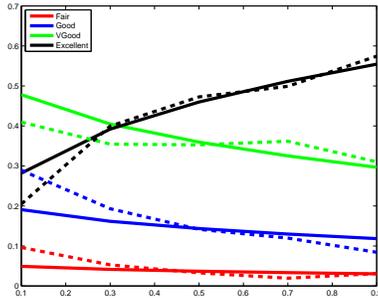


Figure 23: Transition of Excellent Health

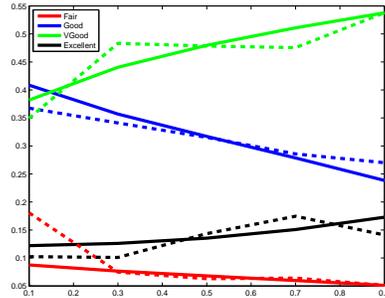


Figure 24: Transition of Very Good Health

Health Shock Probabilities As seen in Table 12, probabilities of getting *any* health shock and a catastrophic health shock (a z -shock) are monotone in health status.

⁶³The PSID has exercise data from 1999 to 2009. The total number of observations for 6 year transition is 13,999.

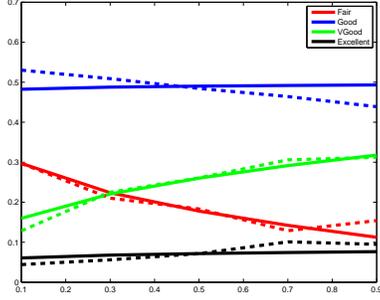


Figure 25: Transition of Good Health

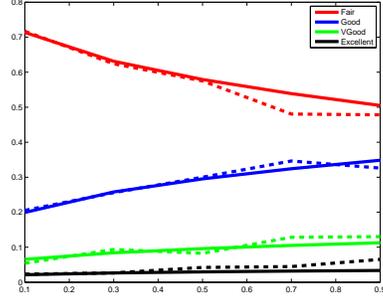


Figure 26: Transition of Fair Health

Table 12: Health Shock Probabilities by Health Status

	Observations	Any Health Shock $1 - g(h)$	z Shock $\kappa(h)$	ε Shock $1 - g(h) - \kappa(h)$
Fair	458	0.78	0.21	0.56
Good	1,139	0.73	0.07	0.67
Very Good	1,618	0.71	0.05	0.67
Excellent	1,143	0.62	0.03	0.59

Effect of Health Shock on Productivity In Table 13, we summarize working hours and labor income reported by those with different health shock categories.

The six year average hours worked of those with z -shocks are about half that of the ones who did not get any shock (and worked) and they earn about half on average. Therefore, we take $\rho = 0.4235$, which is the percentage of labor income earned by those with z -shock, compared to those who have worked and did not experience any health shock (since we denote earnings of those with z -shock as $\rho F(h, 0)$).

Table 13: Hours Worked and Labor Income by Health Shock

		Obs	Mean	Std. Dev.	Min	Max
Hours Worked	All	4,226	1,823	856	0	5,300
	Positive Hours	3,903	1,974	704	7	5,300
	No Shock, Positive Hours	1,259	1,987	781	14	4,732
	z -shock	297	998	1,033	0	3,640
	ε -shock	2,639	1,892	763	0	5,300
Labor Income	All	4,226	30,171	40,573	0	1,153,588
	Positive Hours	3,903	32,362	41,364	0	1,153,588
	No Shock, Positive Hours	1,259	32,606	49,358	0	940,804
	z -shock	297	13,809	25,470	0	253,560
	ε -shock	2,639	31,163	36,883	0	1,153,588

G Additional Quantitative Results

G.1 Model Fit

Figures 27–30 represent the model fit for average effort of each health level.

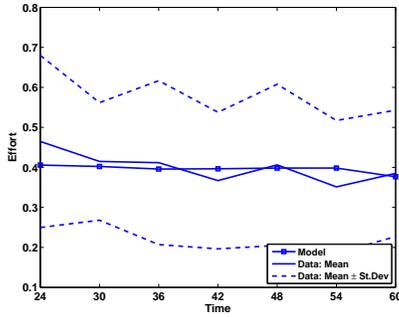


Figure 27: Average Effort: Fair

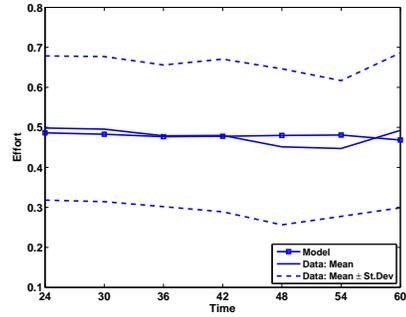


Figure 28: Average Effort: Good

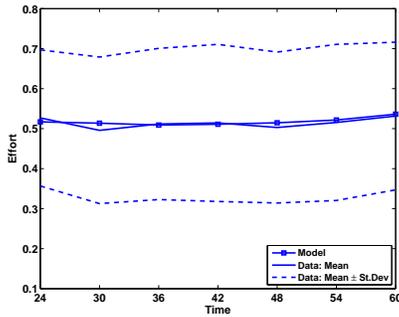


Figure 29: Average Effort: Very Good

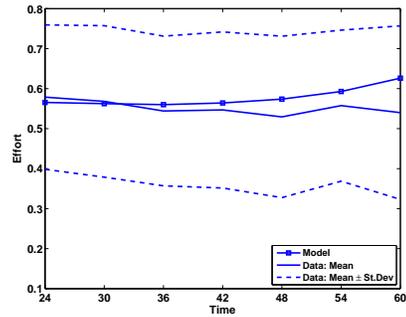


Figure 30: Average Effort: Excellent

G.2 Policy Implications

Insurance Benefits Tables 12 and 13 present the weighted-averages (across education) of the cross-subsidies by health level under different policy regimes. We measure cross subsidies in premium by the differences between the actuarially fair health premium and premium paid under policies; and cross subsidies in wage by the differences between the aggregate wage and productivity of the worker (of a given health level). As discussed in the main text, the negative cross-subsidy implies that the worker is paying higher premium than the actuarially fair price and/or getting paid less in wages than he produces.

Since under no-prior conditions law, only premium is subsidized, and under no-wage discrimination law, only wage is subsidized, we report cross-subsidies of premium and wages under each law. The second row under each health level reports separately the subsidies of premium and wage, under both policies.

Table 14: Cross-Subsidy by Health Level under Different Policy Regimes

Health	Policy	24–29		30–35		36–41	
		Prem.	Wage	Prem.	Wage	Prem.	Wage
Fair	One Policy	0.235	0.269	0.258	0.380	0.249	0.395
	Both Policies	0.145	0.240	0.129	0.341	0.119	0.356
Good	One Policy	0.045	0.108	0.020	0.119	0.001	0.098
	Both Policies	0.022	0.103	0.005	0.107	-0.005	0.083
Very Good	One Policy	-0.030	-0.032	-0.062	-0.118	-0.078	-0.171
	Both Policies	-0.016	-0.029	-0.033	-0.122	-0.044	-0.179
Excellent	One Policy	-0.063	-0.131	-0.091	-0.251	-0.104	-0.315
	Both Policies	-0.037	-0.124	-0.055	-0.252	-0.066	-0.321

Health	Policy	42–47		48–53		54–59		60–65	
		Prem.	Wage	Prem.	Wage	Prem.	Wage	Prem.	Wage
Fair	One Policy	0.265	0.470	0.256	0.461	0.231	0.384	0.174	0.189
	Both Policies	0.117	0.424	0.156	0.418	0.115	0.344	0.116	0.151
Good	One Policy	-0.010	0.101	-0.017	0.089	-0.016	0.068	-0.007	0.024
	Both Policies	-0.009	0.085	-0.011	0.0766	-0.011	0.062	-0.010	0.023
Very Good	One Policy	-0.097	-0.243	-0.100	-0.256	-0.095	-0.215	-0.074	-0.093
	Both Policies	-0.050	-0.253	-0.052	-0.262	-0.052	-0.214	-0.051	-0.082
Excellent	One Policy	-0.123	-0.418	-0.126	-0.431	-0.122	-0.365	-0.106	-0.179
	Both Policies	-0.074	-0.426	-0.076	-0.436	-0.076	-0.362	-0.075	-0.163

Welfare Implications Tables 15 and 16 present the static and dynamic consumption equivalent variations for each education group as well as the aggregates.

	less than HS	HS Grad	Aggregate
Social Planner	5.0301	5.5051	5.4026
Constrained Social Planner	5.0301	5.5051	5.4026
Competitive Equilibrium	0.0000	0.0000	0.0000
No Prior Conditions Law	4.2679	3.6391	3.7748
No Wage Discrimination Law	4.1506	5.2539	5.0157
Both Policies	5.0301	5.5051	5.4026

Table 15: Welfare Comparisons in Static Economy

	less than HS	HS Grad	Aggregate
Social Planner	8.8612	15.6838	14.2113
Constrained Social Planner	6.9535	9.8217	9.2026
Competitive Equilibrium	0.0000	0.0000	0.0000
No Prior Conditions Law	5.6923	6.0978	6.0103
No Wage Discrimination Law	6.3532	9.7194	8.9928
Both Policies	5.4857	9.0633	8.2911

Table 16: Welfare Comparisons in Dynamic Economy

G.3 Sensitivity Analysis with Respect to Ethnicity and Gender

The PSID asks questions on ethnicity⁶⁴, and among them, we take those who answered to be of a national origin (47% of the total sample in 1997) to test robustness. We also restrict our sample to males (about 77%) for the second robustness check.

The health transition function and production function related parameters are the key driving forces of our quantitative results. Therefore, we provide evidence for the similarity in health transition and the labor earnings over the life cycle between the total population and the subsamples.

For the health transition function $Q(h'|h, e)$, we obtain a measure of differences in the estimated probabilities and the data moments, i.e., $\chi^2 = \sum_{i=1, N} \frac{q^{data}(h'^i) - Q^{est}(h'^i)}{Q^{est}(h'^i)}$, where the $q^{data}(h'^i)$ and $Q^{est}(h'^i)$ are the actual data and the estimated probability of a worker with initial health status h with exercise level e ⁶⁵ ending up being health status of h' in the next period. The χ^2 value for the health transition is 1.16 and 1.02 for whites and males, where the $\chi_{49, 0.05}^2$ ⁶⁶ is 79.

With regards to the production function, we provide in Table 17, the data moments associated with the subsamples, in comparison with the full sample. The qualitative features of the moments are similar across different samples: although the absolute numbers for the changes in income over the life-cycle vary in their levels, the gradients over the life cycle are similar. Thus our quantitative results are robust to restricting our samples to white and males.

⁶⁴The exact choices are American (5%); Hyphenated American (e.g., African-American, Mexican-American) (14%); National origin (e.g., French, German, Dutch, Iranian, Scots-Irish) (47%); Nonspecific Hispanic identity (e.g., Chicano, Latino) (2%); Racial (e.g., white or Caucasian, black) (29%) and; 6 Religious (e.g., Jewish, Roman, Catholic, Baptist).

⁶⁵We divide the population into five exercise bins, and use them to evaluate the differences, as we do in our estimation procedure. The only difference is that due to the shortage of observations (since we only use half the total sample), instead of nine bins (in the full model), we use five bins.

⁶⁶The degrees of freedom is 49, as the number of observations are $4 \times 4 \times 5$ (# Health Today \times # Health Tomorrow \times # Exercise Bins), and the number of parameters, 30 (80-1-30). Using the full sample, the χ^2 value is 0.9986.

Moments Description		All	Whites	Male
Income by Age of Less than HS	$t = 1$	-0.0042	-0.0892	0.0320
	$t = 2$	0.1449	0.2026	0.1214
	$t = 3$	0.1715	0.2464	0.1653
	$t = 4$	0.1980	0.2901	0.2091
	$t = 5$	0.0907	0.0014	0.0183
	$t = 6$	-0.0969	-0.3306	-0.1409
	$t = 7$	-0.1112	-0.0970	-0.0742
Income by Age of HS Grad.	$t = 1$	0.2980	0.3019	0.2990
	$t = 2$	0.4738	0.5867	0.4835
	$t = 3$	0.5082	0.6073	0.5522
	$t = 4$	0.5988	0.6274	0.6027
	$t = 5$	0.6060	0.6500	0.6216
	$t = 6$	0.5395	0.5276	0.5366
	$t = 7$	0.2406	0.1792	0.3376
% Income Spent on Med. Exp.	Fair	0.0525	0.0573	0.0482
	Good	0.0429	0.0428	0.0395
	Very Good	0.0353	0.0376	0.0346
	Excellent	0.0308	0.0320	0.0290
% Income Spent on Med. Exp. by Less than HS	Young	0.0386	0.0350	0.0373
	Old	0.0348	0.0357	0.0376
% Income Spent on Med. Exp. by HS Grad	Young	0.0428	0.0465	0.0495
	Old	0.0356	0.0379	0.0447

Table 17: Moments for the Subsample of Population

G.4 Towards Optimal Social Insurance Policy: Detailed Results

Figure 31, plots ex-ante welfare (measured as Dynamic CEV) in the market economy under different tax rates τ and, as point of comparison, also social welfare attained under the different non-discrimination laws as well as under the socially efficient allocation. Table 18 provides the detailed numbers. We observe that the optimal policy a) provides very substantial social insurance and b) comes very close, in welfare terms, to the constrained efficient allocation.

Regime	Dynamic <i>CEV</i>	CV of Con. (Age 42)	CV of Con. (Age 60)	Avg. h (Age 42)	Avg. h (Age 60)
Social Planner	14.2113	0.0000	0.0000	0.0871	0.0845
Const. Social Planner	9.2280	0.1269	0.0691	0.0844	0.0826
$\tau = 0.00$	0.0000	0.5785	0.4526	0.0867	0.0842
$\tau = 0.50$	7.7769	0.2947	0.2299	0.0856	0.0834
$\tau = 0.75$	9.1184	0.1494	0.1161	0.0848	0.0828
$\tau = 0.80$	9.2029	0.1200	0.0932	0.0845	0.0827
$\tau = 0.82$	9.2171	0.1082	0.0840	0.0844	0.0826
$\tau = 0.85$	9.2148	0.0904	0.0701	0.0842	0.0825
$\tau = 1.00$	8.2911	0.0000	0.0000	0.0822	0.0817

Table 18: Incentives vs. Insurance in Tax Equilibria

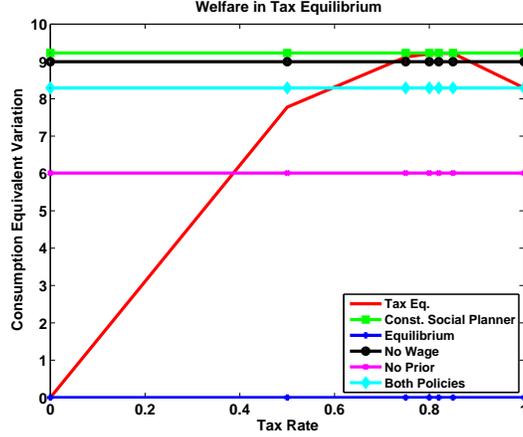


Figure 31: Welfare in Tax Equilibria

H Other Robustness and Sensitivity Analysis: Details

H.1 Insurance Pooling as the Benchmark

Table 19 summarizes the key parameter changes induced by switching the calibrated economy to the partial no prior conditions economy. Table 20 in turn summarizes the new welfare results and contrasts them with the results obtained under perfect competition as the benchmark economy.

	CE as the Benchmark	Insurance Pooling as the Benchmark
$\gamma(h_2)$	0.2086	0.1605
$\gamma(h_3)$	0.0592	0.0431
$\gamma(h_4)$	0.0182	0.0140
ψ	3.1454	3.7884
$v_{T+1}(HS, h_1)$	0.2424	0.1725
$v_{T+1}(HS, h_2)$	1.3846	0.5900
$v_{T+1}(HS, h_3)$	2.1622	1.1362
$v_{T+1}(HS, h_4)$	4.1319	2.1877

Table 19: Parameter Values with Insurance Pooling as the Benchmark

	CE as the Benchmark		Ins. Pooling as the Benchmark	
	$SCEV^i$	$DCEV^i$	$SCEV^i$	$DCEV^i$
Constrained Social Planner	5.4026	9.2026	5.6158	8.7455
Competitive Equilibrium	0.0000	0.0000	0.0000	0.0000
Insurance Pooling Equilibrium	-	-	3.2700	4.5975
No Prior Conditions Law	3.7748	6.0103	3.9027	5.8151
No Wage Discrimination Law	5.0157	8.9928	5.2466	8.5140
Both Policies	5.4026	8.2911	5.6158	7.9891

Table 20: Aggregate Welfare Comparisons with Insurance Pooling as the Benchmark

H.2 Resource Cost and Limited Effectiveness of No Wage Discrimination

Figure 32 divides the parameter space for (γ, τ) into three regions, depending on which policy combination is optimal. We zoom in at that part of the $[0, 1]^2$ -space where the boundaries of the three sets lie.

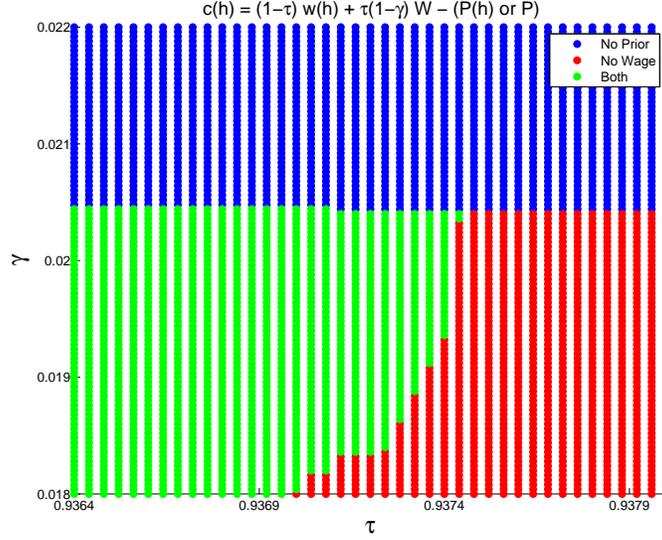


Figure 32: The Best Policy under Limited No Wage Discrimination

H.3 Mismeasured Effort Inputs

For a given household, let the parameter vector be given by $\chi = (\rho, \nu)$, and assume that the cost function is of the form:

$$C(\tilde{e}, e^*, \chi) = \Psi(\chi) [\rho \tilde{e}^\nu + (1 - \rho)(e^*)^\nu]^{\frac{1}{\nu}}.$$

Maximizing $C(\tilde{e}, e^*, \chi)$ subject to equation 43 yields as optimality condition

$$\frac{\tilde{e}}{e^*} = \left[\frac{\lambda(1 - \rho)}{(1 - \lambda)\rho} \right]^{\frac{1}{\nu-1}} := \kappa(\rho, \nu; \lambda). \quad (62)$$

Using this equation in 43 yields as optimal solution $\tilde{e} = \Gamma_1(\rho, \nu; \lambda)e$ and $e^* = \Gamma_2(\rho, \nu; \lambda)e$, where

$$\Gamma_1(\rho, \nu; \lambda) = \frac{\kappa(\rho, \nu; \lambda)}{\lambda \kappa(\rho, \nu; \lambda) + 1 - \lambda} \quad (63)$$

$$\Gamma_2(\rho, \nu; \lambda) = \frac{1}{\lambda \kappa(\rho, \nu; \lambda) + 1 - \lambda} \quad (64)$$

Plugging these solutions into 43 it follows that

$$\begin{aligned} e &= \lambda \tilde{e} + (1 - \lambda) \frac{\Gamma_2(\rho, \nu; \lambda)}{\Gamma_1(\rho, \nu; \lambda)} \tilde{e} \\ &= \eta \lambda \tilde{e}, \end{aligned}$$

where, for a fixed parameter λ , the random variable $\eta = 1 + \frac{(1 - \lambda)\Gamma_2(\rho, \nu; \lambda)}{\lambda \Gamma_1(\rho, \nu; \lambda)}$ has a cross-sectional distribution determined by the population distribution $F(\rho, \nu)$. Finally, the function $\Psi(\chi)$ can be chosen such that $C(\tilde{e}(e), e^*(e), \chi) = e$ and thus $q(e)$ retains the interpretation as utility cost of providing true effort e . This requires

$$\Psi(\chi) [\rho \Gamma_1^\nu + (1 - \rho)\Gamma_2^\nu]^{\frac{1}{\nu}} = 1. \quad (65)$$

Note that under the assumptions on the cost function C , given the observation \tilde{e} and given an assumed distribution for χ , the other aspects of the model are not needed to infer the distribution of e^* . This continues to permit us to estimate Q outside the model despite the fact that \tilde{e}, e^* are endogenous variables. Second, also note that the distribution of e^* is not truncated for any possible observation of \tilde{e} despite the fact that both e, \tilde{e} are restricted to lie in the unit interval. Third, if the distribution of χ and thus of η is degenerate, then \tilde{e} and e^* are perfectly correlated and \tilde{e} becomes a perfect proxy for e . Hence, the extent measurement

error with respect to e given \tilde{e} depends upon the variance of χ and the size of λ . Finally note that the lower is \tilde{e} the smaller is the support of $e = \eta\lambda\tilde{e}$. This implies that high \tilde{e} observations are subject to more doubt than low \tilde{e} observations.⁶⁷ We can therefore rewrite our maximum likelihood problem with noisy effort observations as

$$\max_{\theta} \sum_i \int_{2\lambda-1}^1 \log \left[\frac{Q(h'_i | h_i, \eta\lambda\tilde{e}_i, \theta)}{(2-2\lambda)} \right] d\eta \quad (66)$$

where i indexes the household effort and health status observations. For our application in the main text we set $\lambda = 0.75$. Figure 33 displays our new estimates of Q , with the associated welfare consequences of the policy reforms being reported in the main text.

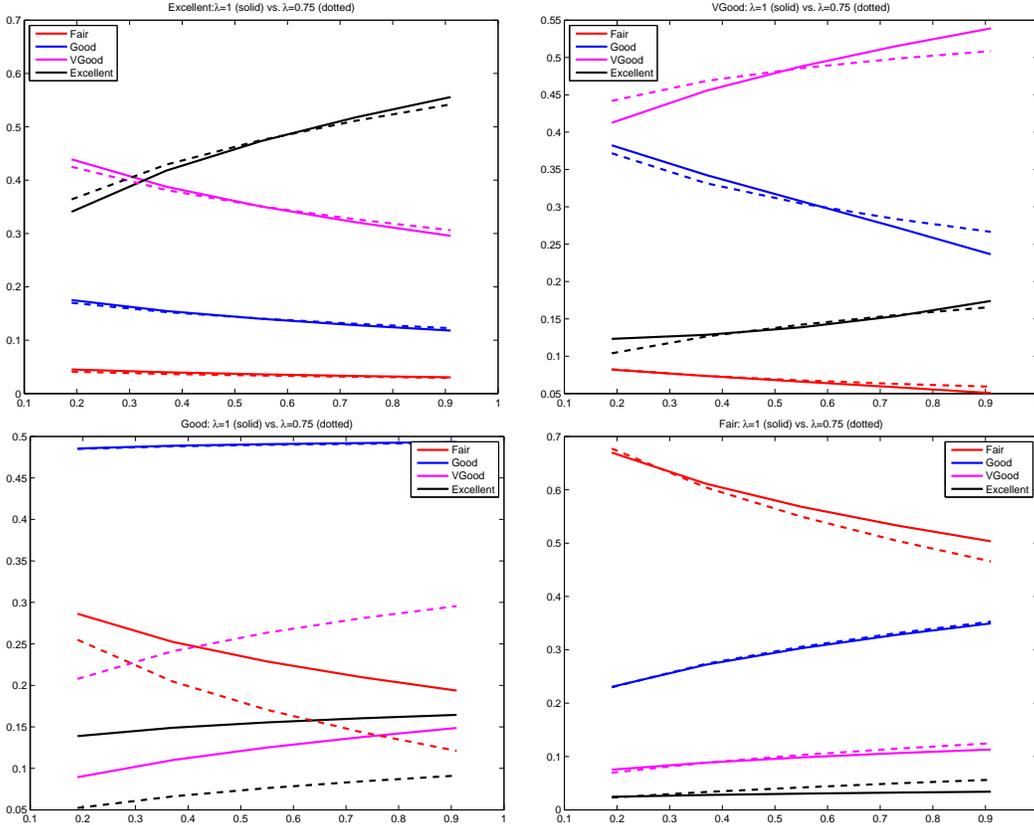


Figure 33: Estimated Health Transition with $\lambda = 0.75$

H.4 Modeling Uninsured Households

Consider a household faced with the wage schedule given in equation 16 and reproduced here for convenience:

$$w(h, \varepsilon - x; \zeta) = \begin{cases} w^{CE}(h; \zeta) & \text{if } \zeta F(h, \varepsilon - x, 0) \geq \zeta F(h, \bar{\varepsilon}^{CE}(h; \zeta)) \\ w^{CE}(h; \zeta) - [\zeta F(h, \bar{\varepsilon}^{CE}(h; \zeta) - \zeta F(h, \varepsilon - x)] & \text{if } \zeta F(h, \varepsilon - x, 0) < \zeta F(h, \bar{\varepsilon}^{CE}(h; \zeta)) \end{cases}$$

We showed in proposition 7 that this wage contract induces a health insurance contract with thresholds $x^{CE}(\varepsilon, h)$, which now is also a function of ζ . Including now the fixed cost for health insurance, expected

⁶⁷Thus measurement error is not classical since

$$E(e|\tilde{e}) = E\{1 + (1 - \lambda)\eta\} \lambda\tilde{e} = \lambda\tilde{e}.$$

utility conditional on purchasing health insurance is given by $U^{CE}(h; \zeta, \Gamma) = u(w(h; \zeta) - P(h; \zeta) - \Gamma)$. If the household does not purchase insurance, but still faces the above wage schedule and can spend on health out of pocket, her associated maximization problem reads as

$$U^{UI}(h; \zeta) = \max_{x(\varepsilon, h; \zeta)} E_{\varepsilon} u [w(h, \varepsilon - x(\varepsilon, h); \zeta) - x(\varepsilon, h; \zeta)]$$

and she would still find it optimal to spend according to the threshold rule with thresholds $x^{CE}(\varepsilon, h; \zeta)$. Now consumption would be stochastic and bear ε -related health expenditure risk, but not wage-risk, given the wage contract and the optimal health expenditure response to it. It is then straightforward to compute, for each health and productivity type (h, ζ) , the fixed cost $\Gamma(h; \zeta)$ that would make a household indifferent between buying and not buying insurance; this threshold simply solves

$$U^{UI}(h; \zeta) = U^{CE}(h; \zeta, \Gamma(h; \zeta)). \tag{67}$$

We plot this measure of willingness to pay in figure 34 below.

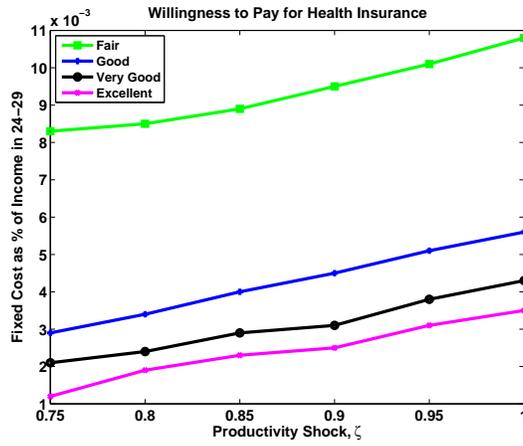


Figure 34: Willingness to Pay for Health Insurance