Fragility: A Quantitative Analysis of the US Health Insurance System*

Björn Brügemann† Yale University

Iourii Manovskii‡ University of Pennsylvania

First version: April 28, 2008

This version: September 25, 2009

Abstract

In this paper we attempt to understand the workings of the US health care system. This system is largely employer based. The dominance of employer provided insurance is typically attributed to its tax deductibility, which is not available to workers purchasing insurance individually. Employers are required to provide the same insurance options at the same cost to all employees. Insurance contracts have a typical duration of one year and insurance companies are allowed to change the rates with few binding restrictions.

One prominent feature of the data is that smaller firms are less likely to provide coverage than larger firms. The reasons for this are unclear. This might be due to the existence of fixed cost of obtaining and maintaining coverage. Alternatively, it is possible that since small firms are less able to pool risk among their workers, making it more difficult for them to provide insurance relative to large firms and forcing them to discontinue coverage occasionally.

We develop a quantitative equilibrium model that features tax deductibility of employer-provided coverage, non-discrimination restrictions, fixed costs of coverage and firms that hire discrete numbers of workers in frictional labor markets. We use the calibrated model to understand what drives the patterns of insurance provision with firm size and to evaluate the effects of this system on the flows of workers across health insurance coverage status, labor market flows, as well as on the the size distribution of firms, and aggregate productivity.

JEL Classification: E20, E24, E25, E62, I10, J32, J63, J78.

Keywords: Health, Health Insurance, Tax Policy, Labor Markets, Labor Mobility, Discrimination.

*This research has been supported by the National Science Foundation Grant No. SES-0617876.

†Department of Economics, Yale University, 28 Hillhouse Avenue, CT 06511. E-mail: bjoern.bruegemann@yale.edu.

‡Department of Economics, University of Pennsylvania, 160 McNeil Building, 3718 Locust Walk, Philadelphia, PA, 19104-6297. E-mail: manovski@econ.upenn.edu.

1 Introduction

Health insurance system for those younger than 65 in the US is largely employer-based. More than three-fifths of Americans under age 65 are covered by employer-sponsored health plans. Among those with private coverage from any source, 94.6 percent of adults and 94.0 percent of children held employment-related coverage in 2006 (Selden and Gray (2006)). The dominance of employer provided insurance is typically attributed to its tax deductibility, which is not available to workers purchasing insurance individually.[1] Despite the economic magnitude of this system, its properties are not well understood. In this paper we attempt to fill some of the gaps in our understanding of how this system works and of some of the economic implications of is design.

It is well documented that large firms are much more likely to offer health insurance. Only 51.8 percent of establishments in firms of fewer than 50 workers offer health insurance to any of their workers, compared with 88.7 percent of establishments in firms of 50 or more workers[2] What might cause such a difference? First, this may be due to fixed costs of offering health insurance. Small firms have fewer workers over which to spread these fixed costs. Second, it may be due to the limited ability of small firms to pool risk. We label this a "fragility" channel. Loosely speaking, if an employee in a small firm becomes ill, maintaining insurance would imply a large increase in premiums for all workers. The tax subsidy may then be insufficient to make employer provided insurance attractive to healthy workers, who can switch to individual insurance, so that employer provided insurance "unravels." The relative importance of these two (and possibly other) channels in explaining the observed variation of insurance provision with firm size is not

[1]Under current law, employer-provided health insurance coverage is excluded from employees' income for determining their federal income taxes. Exclusions also apply to Social Security, Medicare, and unemployment taxes (both employer and employee portions) and to state income and payroll taxes as well. Considering the average cost of employment-based insurance, now around $4,750 a year for single coverage and $12,700 for family coverage (Lyke (2008)), these exclusions result in significant costs to the government. Joint Committee on Taxation (2008) estimated calendar year 2007 tax expenditures for the employer coverage exclusion to have been between $105 and $145.3 billion for the federal income tax and $100.7 billion for Social Security and Medicare taxes. The federal income tax component alone represents the single largest source of revenue loss in the U.S. Budget and is, e.g., 60 percent larger than the revenue loss from the federal income tax deduction of mortgage interest and 3 times larger than the current revenue loss form the tax-deductibility of contributions and earnings in 401(k) retirement plans (Table 19-1 in Office of Management and Budget (2008)).

[2] See (2) for a description of the data on which these estimates are based.

well understood. While fixed costs are easy to model, we are not aware of any existing model that allows to evaluate the importance of the fragility channel. We build such a model in this paper.

The key but largely ignored in the literature feature of the US health insurance system is the non-discrimination restrictions: by law employers are required to provide the same insurance options at the same cost to all employees.[3] In addition to not being able to discriminate on the basis of health in compensation, employers are not able to discriminate on the basis of health in hiring or firing. If employer had a tax subsidy for providing insurance but were able to either negotiate individual contracts for its workers, or to discriminate in pay based on individual worker's health expenses, there would be no pooling of risk across workers. Each worker's premiums would equal his expected costs plus a load equal to the cost of administering insurance. Thus, a firm would purchase insurance individually for each worker, pay for it by reducing the worker's wage by the cost of this insurance, and split the tax benefit with the worker according to the wage setting protocol (e.g., bargaining). In this case tax subsidy for firm provided insurance would simply represent a transfer to firms which does not appear to have been the intention of the policymakers.[4] Thus, we explicitly introduce non-discrimination regulation into the model.

To study the fragility channel, we cannot assume that firms have a continuum of workers (of possibly different measures). Thus, we work in an economy where workers are discrete. This makes the model somewhat complicated to compute but enables us to directly measure the quantitative importance of the fragility channel.

There are several additional questions that we would like to address with our model.

1. We would use the model to understand how effective the current system is in providing coverage to workers. In particular, we would like to know how far the current

[3]The nondiscrimination provisions of Health Insurance Portability and Accountability Act of 1996 (HIPAA) prohibit insurers or employers from excluding, providing less coverage to, or charging higher premiums to any individual in a group due to his or her health status, but the entire group could be charged a higher premium (or have benefit exclusions). Also, HIPAA requires insurers to offer insurance to any small employer, but does not restrict the premiums charged. Most states have laws restricting in various ways the ability of insurers to price discriminate across groups.

[4]Thomasson (2002, 2003) and Lyke (2008) describe the historical evolution of the US health insurance system.

system that features the interaction of tax subsidy and non-discrimination restrictions is from the national health insurance (which would arise if the model featured just one large firm). Ultimately, we would like to use the model as a laboratory to evaluate various reform proposals.

2. It has been noted in the literature that the employer based health insurance system must affect the labor market flows (see Dey and Flinn (2005) for a review of (contradictory) findings in this literature). If the fragility channel or the fixed costs of setting up the coverage make it relatively more difficult for small firms to provide coverage, employer-based health insurance must also have non-trivial effects on the size distribution of firms and industry dynamics. How important are these effects? What are the true costs (or benefits) of the employer-based health insurance system when the consequences of these distortions are measured? Guner, Ventura, and Yi (2008) and Restuccia and Rogerson (2008) provide evidence that firm-size dependent policies may have very large aggregate effects.

3. In 2007 19.7 percent of adults between the ages of 19 and 65 were not covered by health insurance (Henry J. Kaiser Family Foundation (1998)).[5] However, this is not a static pool of people, there are substantial flows in and out of it. One of the objectives of this paper is to develop a model that may shed light on the reasons for these flows. We expect that the fragility of coverage by small firm plus worker flows across employment and non-employment states and across firms that do and do not offer health insurance may play a nontrivial role in accounting for these flows.

Two other features of the tax-treatment of the US employer-based health insurance system have received substantial attention in the literature. First, since at least Feldstein (1973), it has been recognized that since health insurance is subsidized workers may demand too much of it leading to excessive insurance coverage and costs. Second, it is argued to be regressive because workers' tax savings depend on their marginal tax rates. Since marginal tax rates generally increase with income, higher income individuals and families obtain greater tax savings. Jeske and Kitao (2008) study quantitatively the

[5]Interestingly, over 80 percent of the uninsured are wage earners or members of working families.

welfare consequences of the regressive tax treatment of health insurance. For now, at least, we do not study these trade-offs.

Dey and Flinn (2005) present an equilibrium model of health insurance provision by firms and wage determination. They investigate the effect of employer-provided health insurance on job mobility rates and economic welfare using an on-the-job search and bargaining framework. They find that the employer-provided health insurance system does not lead to any serious inefficiencies in mobility decisions. However, they assume that wages and health coverage are negotiated between workers and firms on an individualistic basis, that is, without reference to the composition of the firm's current workforce. Moreover, they do not model the favorable tax treatment of the employer-provided health insurance coverage. Thus, they abstract from the key mechanisms evaluated in this paper.

Jeske and Kitao (2008) also study the effects of tax deductibility of group coverage using a dynamic general equilibrium model with heterogeneous agents. The key difference is that they do not model firms. Instead, it is assumed that workers either have a stochastic opportunity to join a group plan (or opt out) or they do not have such an opportunity and have a choice of whether to purchase individual coverage. Thus, all workers with the opportunity to purchase health insurance belong to one large group. In difference, in our model, such groups are defined at the firm level. As in our model, healthy workers that participate in the group health plan subsidize unhealthy workers. The extent of their willingness to do so depends on the amount of tax savings they obtain. Without the tax subsidy, group insurance unravels. The model in Jeske and Kitao (2008) is not designed to address the fragility of coverage in small firms, the effects of the health insurance system on labor market flows, and its heterogeneous effects on firms of different sizes.

The paper is organized as follows. In Section 2 we describe the key facts motivating our analysis. In Section 3 we develop the model of the employer-provided health insurance coverage that features the fragility channel. In Section 4 we define equilibrium. In Section 5 we calibrate the model and perform the quantitative analysis of the current system and evaluate the effects of several proposed reforms. Section 6 concludes.

2 Facts

2.1 Data Sources

2.1.1 1997 Robert Wood Johnson Foundation Employer Health Insurance Survey

The 1997 Robert Wood Johnson Foundation Employer Health Insurance Survey (Long and Marquis (1997)) is a nationally representative survey of public and private employers (with at least one employee) conducted in 1996 and 1997. Data were collected on employers' offers of health insurance coverage, employees' eligibility and enrollment in health plans, and, for each plan offered, the plan type (HMO, POS, PPO, conventional), premiums (employer and employee contributions), benefits, cost-sharing, and employer self-insurance status. The study also collected information on the characteristics of employers and workers, including the number of employees at the establishment, the number of employees statewide and nationwide, and the distribution of workers by hours worked, age, sex, and earnings.

Our analysis is based on a sample of $21,545$ private sector employers only. All results are weighted by the sampling weights provided by the Survey.

2.2 Insurance Provision by Firm Size

It is well-known that insurance provision varies substantially with firm size. In particular, Figure 1 plots the fraction of firms providing insurance as a function of firm size. Overall, approximately 50% of all employers provide health insurance to their workers. However, only 31% of employers with less than 10 employees provide coverage. Insurance provision increases rapidly with firm size to 62% of employers with 10 to 24 employees, 73% of employers with 25 to 49 employees, and 89% of employers with more than 50 employees.

Figure 2 illustrates that this pattern is robust to controlling for the average wage received by workers. Firms that pay higher wages are indeed more likely to provide health insurance but the gradient of the probability of coverage with respect to firm size is almost independent of the average payroll. Strikingly, the lowest paying firms with more than 50 employees are 60% more likely to offer health insurance than the highest paying firms with less than 10 employees.

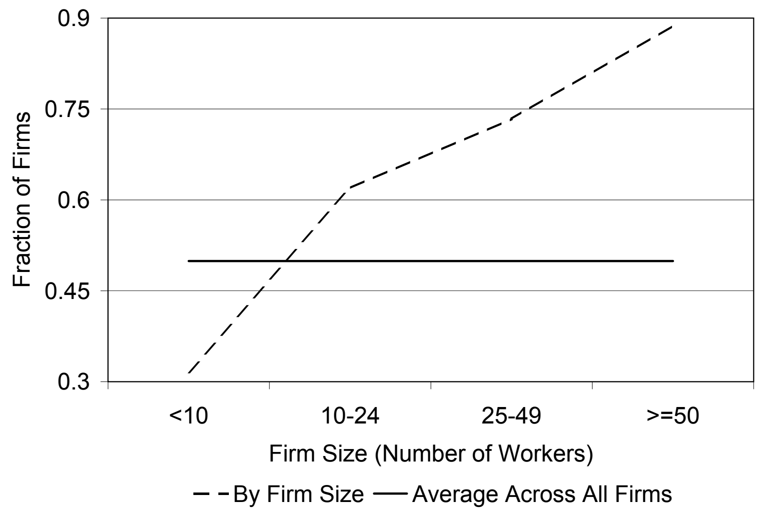

Figure 1: Fraction of Firms Providing Insurance by Firm Size.

2.3 Discontinuance of Insurance Provision by Firm Size

One issue that has received little attention in the literature (Long and Marquis (1998) is one exception) is that firms occasionally discontinue offering insurance coverage to their workers. The probability of discontinuing coverage varies systematically with firm size. In particular, Figure 3 illustrates that 13% of firms that offered insurance within two years prior to the survey date were not offering at the time of the survey. This fraction is over 21% for firms with less than 10 employees and declines to just 4% of firms with more than 50 employees.[6]

2.4 Variability of Insurance Premiums by Firm Size

Health insurance premiums faced by firms are quite volatile over time. Moreover, this volatility is also systematically related to firm size.

The average increase in premium in the RWJS sample of private sector firms providing coverage in 1996 and 1997 was around 2.5% irrespective of the firm size. This relatively

[6] May also mention here the opposite flow documented in Long and Marquis (1998). They find that almost one-fifths of the firms with less than 10 employees that did not offer insurance in 1995 offered it in 1997, in contrast to only 6 percent of firms with more than 100 employees making such change.

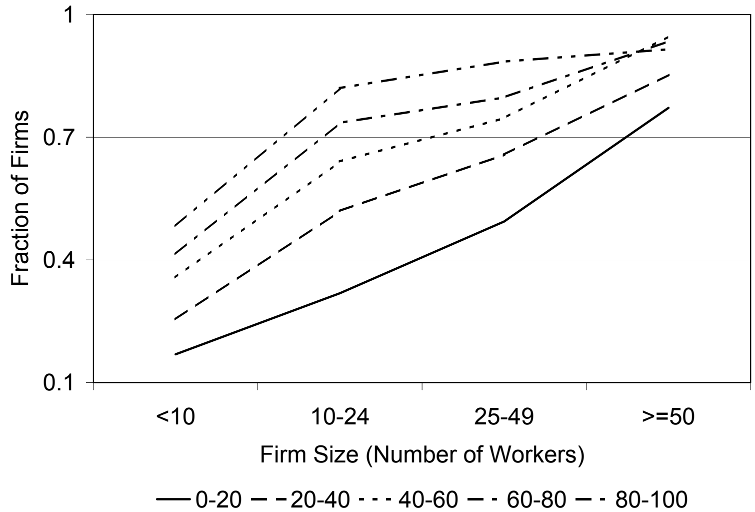

Figure 2: Fraction of Firms Providing Insurance by Firm Size and Average Payroll.

small increase in private health insurance premiums in 1997 accords well with other sources of data.[7]

The average annual absolute percentage change in premiums across firms of all sizes was considerably higher at 6%. This indicates that firms experience substantial changes (positive and negative) to premiums from one year to the next. The average annual absolute pecentage change in premiums declines from 7.2% for firms employing less than 10 workers to 4.8% for firms with more than 50 employees. As an alternative statistic, consider the standard deviation of the premium change for firms of different sizes. It equals 9.9 for firms of all sizes and declines from 11.5 for firms with less than 10 employees to 10.3, 8.9, and 8.0 for firms with 10-24, 25-49, and more than 50 employees, respectively. Finally, consider the difference between the 90th and 10th percentile of the premium change distribution. This statistic equals 18% for firms of all sizes and declines from 24% for firms with less than 10 employees to 19%, 17%, and 14% for firms with 10-24, 25-49, and more than 50 employees, respectively.[8]

[7]E.g., The U.S. Department of Health and Human Services Fact Sheet available at http://www.hhs.gov/news/press/2000pres/20000110.html.

[8]Cutler (1994) finds qualitatively similar patterns in the 1991 Health Insurance Association of America

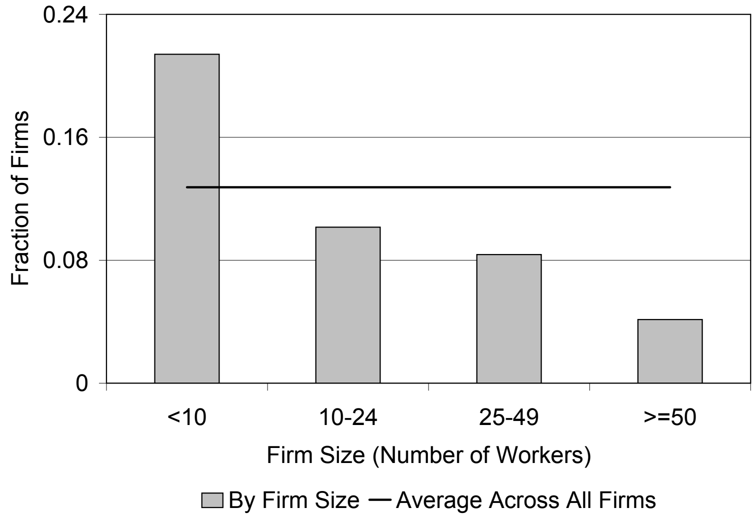

Figure 3: Fraction of Firms Discontinuing Insurance over a 2 Year Period by Firm Size.

2.5 Features of Observed Insurance Contracts

In practice health insurance contracts have a typical length of one year. Long-term health insurance contracts are virtually non-existent. Moreover, premiums adjust almost freely upon renewal (subject to some restrictions imposed in several states). Insurance premiums are typically based at least in part on the expected medical costs of the firm purchasing insurance, a method termed "experience rating" in the literature. Thus, firms that experience adverse events will pay more for insurance than firms that do not.

These features imply that the typical insurance contract insures the firm against the risk that the medical costs within a year will exceed their expected value at the beginning of the year. However, if some workers in a firm learn that they are permanently

(HIAA) survey. He finds that the spread between the 90th and 10th percentile of the costs of comparable plans is 174% for firms with leass than 50 workers, and it declines monotonically to 71% for firms with 501 to 1000 workers. The spread between the 90th and 10th percentile of the change in costs from one year to the next is 45% for firms with leass than 50 workers, and it declines monotonically to 23% for firms with 501 to 1000 workers. Thus, he finds a larger spread of the distribution of cost changes. The difference may be attributable to the difference in survey years (1991 being a recession year and exhibiting insurance premium increase of 14% compared to 2.5% in 1997). Another difference between our studies is that we estimate the percent change in health insurance cost between the two years as directly reported by the respondent. Cutler's study is based on the data about the actual premiums paid which he attempts to adjusts for the different generosity of the plans.

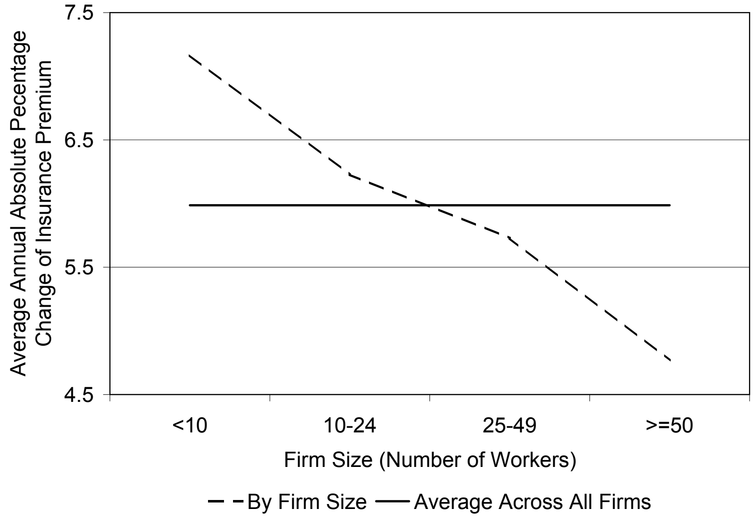

Figure 4: Variability of Insurance Premiums.

less healthy than they thought[9], future premiums will be higher than they expected. Current system does not provide insurance against such intertemporal risk, except to the extent that large firms can achieve a relatively stable overall health composition of their workforce.

There is no consensus in the literature as to what forces have shaped the system to have these features. Some of the hypothesis offered in the literature include the presence of adverse selection and moral hazard, whereby had the insurers not experience rated, they would attract firms with higher expected health costs or encourage less healthy behavior on the part of those insured. Another possibility involves inability of firms and insurers to commit to long-term contracts. If firms can walk away from a contract, the firms that have learned that their workers are healthier than they thought will do so. Firms might be reluctant to agree to sizable pre-payment required to overcome this selection problem given the uncertainties they face about the future. Cochrane (1995) suggests that the competition for healthy groups among health insurers is a relatively

[9]For example, having a heart attack today signals a greater risk of a heart attack in the future. Babies born with birth defects often require medical care throughout their life.

recent phenomenon to which the regulatory and legal systems have not adapted yet to enforce potentially feasible long-term contracts. In any case, we do not think this paper is a place to engage in this debate. Instead, we assume that the nature of the contracts offered is a feature of the environment that we take as given.

2.6 Legal Framework

No Mandate

- Firms are not required to offer health insurance.

Tax Deductibility

- Employer purchase of health insurance is fully tax deductible. Individual purchase is not.

Non-Discrimination

- Health plans cannot discriminate in access or cost of coverage based on health.

3 Model

At the beginning of every period there is a mass one of workers. Workers permanently leave the labor market with probability $1-\rho$ per period and leavers are replaced by labor market entrants. There is a large mass of potential firms.

Preferences of workers are

$$\mathbb{E}_t \sum_{t'=t}^{\infty} \beta^{t'-t} U(c_{t'})$$

with $U' > 0$ and $U'' \leq 0$. Firms share the discount factor β but are risk neutral.

Firms are subject to idiosyncratic productivity shocks: z indexes the productivity of a firm, which follows a discrete Markov process with transition probabilities $q^Z_{zz'}$.

Output of a firm with productivity z is given by

$$zF(g_e),$$

where g_e is the number of workers and $F' > 0$, $F'' \leq 0$. Active firms are subject to a fixed operating cost c_f, which can only be avoided through exit. New firms can enter by paying an entry cost c_e, and draw an initial productivity from the invariant distribution associated with q^Z.

Workers are subject to idiosyncratic shocks to their health. Their health status follows a two state Markov process: healthy h or unhealthy u with transition probabilities $q^H_{ii'}$ for $i, i' \in \{h, u\}$. We assume persistence $q^H_{hh} > q^H_{uh}$. Health status in turn determines medical expenditures $e_u > e_h$. Let q^0_i denote the fraction of labor market entrants in health status i.

The labor market is not competitive due to search frictions, to be described below. Compensation consists of wage payments and health insurance. Due to regulation, compensation cannot discriminate between healthy and unhealthy workers. To keep the determination of compensation as simple as possible given this restriction, we assume that firms have all the bargaining power.

Non-discrimination regulations also apply to dismissal. To capture these constraint, we model the dismissal decision of a firm in two steps, with the following timing. First, the firm decides how many workers of each health status to retain. In this step, the firm is subject to the constraint that workers asked to leave must not prefer to stay, given the compensation package. For example, if a firm offers health insurance and unhealthy workers prefer to stay, the firm cannot dismiss them in this step. In the second step, the firm can dismiss workers without considering health status, not facing the constraint that workers must not prefer to stay. In addition to endogenous dismissal, an employed worker separates exogenously with probability δ every period.

A firm can recruit new workers by posting vacancies. It posts $g_v \in \{0, 1, \ldots\}$ vacancies at cost c_v per vacancy. We assume that there is an upper bound on total firm employment $\bar{\psi}$.

The probability that a vacancy contacts a worker and the probability that a searching worker contacts a vacancy are given by

$$q(\theta) = M\left(\frac{1}{\theta}, 1\right) \quad \text{and} \quad f(\theta) = M\left(1, \theta\right), \tag{1}$$

respectively. Here M is a constant returns to scale matching function and $\theta = \frac{v}{m}$ is the ratio of vacancies to searchers.

Firms can purchase health insurance contracts that cover the medical expenditures of their employees. While wage income is taxed at rate $\tau \in [0, 1)$, the provision of health insurance is not subject to taxes. Health insurance carriers charge an administrative load which may depend on the number of employees $\kappa(g_e)$. Apart from these administrative costs insurance is offered on actuarially fair terms.

We assume that if a firm begins health insurance coverage, this coverage expires stochastically with probability q^I every period, upon which it decides upon renewal. This allows for the possibility that the frequency of decisions about health insurance coverage is lower than the frequency of decisions concerning dismissal and recruitment. A firm can then be in one of three insurance provision states, which we collect in the set $\{C, E, N\}$: C indicate a continuing insurance spell, that is the firm provided coverage in the previous period and the contract has not expired; E indicates that the firm provided coverage in the preceding period but the contract just expired; N indicates no coverage in the previous period. We allow for size dependent costs of beginning and discontinuing health insurance coverage, denoted $b_I(g_e)$, $d_I(g_e)$, respectively. Workers who do not receive coverage through their employer pay out of pocket for their medical expenditures.

Events within a period unfold as follows. New firms enter. Firms decide on exit, health insurance, wages, retainment and recruitment of workers. Production takes place. Health shocks are realized and insurance payments are made. Consumption takes place. Some workers exit the labor market and some employed workers separate exogenously. Idiosyncratic shocks to firm productivity are realized. Searching workers are matched with vacancies.

4 Stationary Equilibrium

The state of a firm at the beginning of the period $s = (z, \psi_h, \psi_u, I)$ is given by its productivity $z \in \mathcal{Z}$, its workforce $(\psi_h, \psi_u) \in \Psi$, and its insurance coverage status $I \in \{C, E, N\}$. Let $\mathcal{S} = \mathcal{Z} \times \Psi \times \{C, E, N\}$ denote the state space. For $s \in \mathcal{S}$ we write $z(s)$ for firm productivity in that state, using analogous notation for the other state variables. In a given period a firm make seven choices, which are collected in a policy vector $g = (g_h, g_u, g_e, g_v, g_I, g_w, g_x)$. The first three entries concern dismissal of incumbent employees. In the first dismissal step, the firm chooses the number of healthy workers g_h and unhealthy workers g_u to retain. In the second step, the firm dismisses workers at random, and g_e denotes the total number of workers retained after random dismissal. The fourth choice concerning the size of the workforce is the number of vacancies g_v. The next two choices describe the compensation package offered by the firm. Whether to offer insurance is a binary choice $g_I \in \{0, 1\}$. The decision concerning wages is binary as well: since the firm has all the bargaining power it always makes one of the two worker types indifferent between staying and leaving, hence the wage decision reduces to the decision which type to make indifferent $g_w \in \{h, u\}$. The final entry is the decision of the firm whether to exit $g_x \in \{0, 1\}$.

Let $\mathcal{G}$ denote the set of all pure policies which are consistent with the upper bound on firm size

$$\mathcal{G} \equiv \{(g_h, g_u, g_e, g_v, g_I, g_w, g_x) \in \mathbb{N}_0^4 \times \{h, u\} \times \{0, 1\}^2 | g_e + g_v \leq \bar{\phi}\}.$$

The equilibrium concept allows firms to use mixed policies, i.e. each firm chooses an element of $\Gamma \equiv \Delta(\mathcal{G})$, the set of all probability distributions over the set $\mathcal{G}$. Having defined the state space and the policy space, next we provide a list of objects that make up a stationary equilibrium. A stationary equilibrium consists of:

1. A firm value function $J(\cdot) : \mathcal{S} \to \mathbb{R}$, with $J(s)$ giving the value of a firm in state s.

2. A firm policy function $\gamma(\cdot|\cdot) : \mathcal{S} \to \Gamma$, where $\gamma(\cdot|s) \in \Gamma$ is the mixed policy for firms in state s, Γ is the set of all mixed policies, and $\gamma(g|s)$ is the probability that a firm in state s implements pure policy $g \in \mathcal{G}$.

3. A wage function $w(\cdot, \cdot) : \mathcal{G} \times \mathcal{S} \to \mathbb{R}$ with $w(s, g)$ giving the wage a firm in state s would pay if it were to adopt policy g.

4. A correspondence $\mathcal{G}(s) \subset \mathcal{G}$ for $s \in \mathcal{S}$, where $\mathcal{G}(s)$ is the set of policies feasible for a firm in state s.

5. Worker value functions $V_h(\cdot), V_u(\cdot) : \mathcal{S} \to \mathbb{R}$, with $V_i(s)$ giving the utility of a worker with health status $i \in \{h, u\}$ employed in a firm in state s.

6. Worker continuation values $C_h(\cdot,\cdot), C_u(\cdot,\cdot) : \mathcal{S} \times \mathcal{G} \to \mathbb{R}$ giving the continuation value of a worker in a firm in state s pursuing policy g.

7. Values of searching $V_h^s, V_u^s \in \mathbb{R}$, for healthy and unhealthy workers, respectively.

8. An invariant distribution $\mu(\cdot) : \mathcal{S} \to [0,1]$ of the state of firms at the beginning of the period.

9. A mass of workers $m \in [0,1]$ searching for employment, and the fraction of searching workers $\nu \in [0,1]$ which is healthy.

10. A mass of active firms ϕ.

11. Labor market tightness $\theta \in \mathbb{R}_+$.

In the following subsections we derive the conditions relating these objects, followed by a formal definition of stationary equilibrium.

4.1 Firm Decision

Consider a firm in state $s = (z, \psi_h, \psi_u, I) \in \mathcal{S}$. The expected flow utility of a worker that remains with the firm at the time of production, as a function of wage w and insurance provision g_I, is given by

$$\begin{aligned} u_i(w, g_I) &= g_I U((1-\tau)w) \\ &\quad + (1-g_I)\Big[q_{ih}^H U((1-\tau)w - e_h) + q_{iu}^H U((1-\tau)w - e_u)\Big]. \end{aligned}$$

The first term is the flow utility of the worker if insurance is provided by the employer. The second term gives expected flow utility of the worker if health expenditures must be paid out of pocket.

The lifetime expected utility at the time of production of a worker with health status i is given by $u_i(w(s,g), g_I) + \beta\rho C_i(s,g)$. This must be equal to $V_{g_w}^s$ for type g_w if the firm chooses to make type g_w indifferent. Thus the following equilibrium relationship implicitly determines the wage $w(s,g)$:

$$u_{g_w}(w(s,g), g_I) + \beta\rho C_{g_w}(s,g) = V^s_{g_w} \tag{2}$$

Next we determine the set of pure policies that are feasible for a firm in state s, denoted $\mathcal{G}(s) \subset \mathcal{G}$. For g to be feasible it must be that workers asked to leave must not prefer to stay. The choice of the wage insures that this is true for type g_w. For the other type $\neg g_w$ it must be that

$$u_{\neg g_w}(w(s,g), g_I) + \beta\rho C_{\neg g_w}(s,g) \geq V^s_{\neg g_w} \quad \text{if} \quad g_{\neg g_w} > 0. \tag{3}$$

However, if $g_i < \psi_i(s)$, then the pure policy g also calls on some workers of type i to leave. A leaving worker can induce a deviation from the firm policy by staying. Let $G^+_i(g)$ be the same pure policy as g except that one additional worker of type i stays:

$$G^+_h(g_h, g_u, g_e, g_v, g_I, g_w, g_x) \equiv (g_h + 1, g_u, g_e + 1, g_v, g_I, g_w, g_x)$$

for healthy workers, with G^+_u defined analogously. For g to be feasible it must be that

$$u_i(w(s,g), g_I) + \beta\rho C_i(s, G^+_i(g)) \leq V^s_i \quad \text{if} \quad g_i < \psi_i(s). \tag{4}$$

Finally, if a firm is in a continuing insurance spell it must provide coverage

$$g \in \mathcal{G}(z, \psi_h, \psi_u, C) \Rightarrow g_I = 1. \tag{5}$$

Thus the set of feasible policies for a firm in state s is

$$\mathcal{G}(s) = \{g \in \mathcal{G} | g_h \leq \psi_h \wedge g_u \leq \psi_u \wedge \text{(3)–(5) hold}\}. \tag{6}$$

It is convenient to define net revenue for active firms, deducting fixed cost of operating, health insurance premiums, recruiting costs, and applicable costs of beginning or discontinuing health insurance coverage from output:

$$\begin{aligned} R(s,g) \equiv\ & z(s)F(g_e) - c_f - g_I p(g) g_e - c_v g_v \\ & - b_I(g_e) g_I \mathcal{I}(I(s) = N) - d_I(g_e)(1 - g_I)\mathcal{I}(I(s) = E). \end{aligned}$$

Here premium of employer provided insurance $p(g)$ is given by

$$p(g) \equiv (1+\kappa(g_e))\frac{g_h(q^H_{hh}e_h + q^H_{hu}e_u) + g_u(q^H_{uh}e_h + q^H_{uu}e_u)}{g_h + g_u},$$

where $\kappa(g_e)$ is the administrative load.[10] The current state s, a pure policy g, tightness θ, and the fraction of healthy workers among searchers ν, together induce a distribution over the firm's future state $\mu(s'|s, g, \theta, \nu)$, which is derived in Appendix I.1. The firm value function $J(\cdot)$ must satisfy the Bellman equation

$$J(s) = \max_{g \in \mathcal{G}(s)} g_x \left[R(s,g) - w(s,g) + \beta \sum_{s' \in \mathcal{S}} J(s')\mu(s'|s,g,\theta,\nu) \right] \quad \text{for all} \quad s \in \mathcal{S}. \tag{7}$$

The policy function $\gamma(\cdot|\cdot)$ must satisfy $\gamma(\cdot|s) \in \Gamma$ for all $sin\mathcal{S}$ and

$$\gamma(g|s) > 0 \quad \Rightarrow g \in \arg\max_{g \in \mathcal{G}(s)} g_x \left[R(s,g) - w(s,g) + \beta \sum_{s' \in \mathcal{S}} J(s')\mu(s'|s,g,\theta,\nu) \right] \tag{8}$$

for all $g \in \mathcal{G}$ and all $s \in \mathcal{S}$.

4.2 Worker Value Functions and Continuation Values

The probability that a worker in health status i is retained if its employer is in state s and pursues policy g is

$$\sigma_i(s,g) \equiv \frac{g_i}{\psi_i(s)} \frac{g_e}{g_h + g_u}.$$

where the first factor is the probability of being retained during selective dismissal, and the second factor is the probability of being retained during random dismissal. The worker value functions must then satisfy the equilibrium relationship

$$V_i(s) = \sigma_i(s,g)\left\{u_i(w(s,g), g_I) + \beta\rho C_i(s,g)\right\} + \left[1 - \sigma_i(s,g)\right] V_i^s. \tag{9}$$

The current state s, a pure policy g, tightness θ, and the fraction of healthy workers among searchers ν, together induce a distribution over the firm's future state denoted

[10]The formula implicitly assumes that premiums are paid before random dismissal. Thus $p(g)$ is based on the expected health expenditures given the health composition of the workforce before random dismissal. This is actuarially fair, since in expectation random dismissal does not change the health composition of the workforce. Since firms are risk neutral, nothing would change if premiums are based on the health composition after random dismissal, since expected premium cost associated with providing coverage are the same.

$\mu_{ii'}\left(s' \left| s, g, \theta, \nu \right.\right)$ where the subscripts indicate that this distribution is conditional on the worker transiting from health status i to i'. This distribution is derived in Appendix A.2. Worker continuation values must satisfy the equilibrium relationship

$$
\begin{aligned}
C_i(s,g) = (1-\delta)\Big\{ & q^H_{ih} \sum V_h(s')\mu_{ih}\left(s' \left| s, g, \theta, \nu \right.\right) \\
& + q^H_{iu} \sum V_u(s')\mu_{iu}\left(s' \left| s, g, \theta, \nu \right.\right) \Big\} + \delta \left[q^H_{ih} V^s_h + q^H_{iu} V^s_u \right]
\end{aligned}
\tag{10}
$$

4.3 Value of Searching

The flow utility of a searching worker with health status i is

$$
u^s_i = q^H_{ih} U(b - e_h) + q^H_{iu} U(b - e_u),
$$

where b is the flow value of non-market activity, and searchers pay health expenditures out of pocket. The value of searching of a worker in health status i must satisfy the equilibrium relationship

$$
\begin{aligned}
V^s_i = {} & u^s_i + \beta\rho\left(1 - f(\theta)\right)\left[q^H_{ih} V^s_h + q^H_{iu} V^s_u \right] \\
& + \beta\rho f(\theta) \Big\{ q^H_{ih} \sum V_h(s')\mu^s_h\left[s' \left| \mu(\cdot), \gamma(\cdot|\cdot), \theta, \nu \right.\right] + q^H_{iu} \sum V_u(s')\mu^s_u\left[s' \left| \mu(\cdot), \gamma(\cdot|\cdot), \theta, \nu \right.\right] \Big\}
\end{aligned}
\tag{11}
$$

where the first part of the continuation utility corresponds to contacting a firm, while the second part applies in the absence of a contact. Here $\mu^s_i\left(s' \left| \mu(\cdot), \gamma(\cdot|\cdot), \theta, \nu \right.\right)$ for $i \in \{h, u\}$ is the the distribution of the state of the worker's new firm conditional on the worker having new health status i, and it is derived in Appendix I.3.

4.4 Entry of New Firms

Let μ^Z denote the invariant distribution induced by the transition probabilities $q^Z_{zz'}$. The entry condition is

$$
c_e = \sum_{z \in \mathcal{Z}} J(z, 0, 0, N)\mu^Z(z), \tag{12}
$$

where the initial state of a new firm has zero employment and no past insurance coverage.

4.5 Invariant Distribution of Firms

In a stationary equilibrium the number of firms must remain constant, thus entry must compensate for exit. This implies that we can compute the invariant distribution by taking the transition law $\mu(s'|s, g, \theta, \nu)$, modifying it by letting exiting firms start over as new entrants. Using superscript x to denote application of the operator that performs this modification, the invariant distribution $\mu(s)$ must satisfy the equilibrium relationship

$$\mu(s') = \sum_{s \in \mathcal{S}} \sum_{g \in \mathcal{G}(s)} \mu^x(s'|s, g, \theta, \nu)\gamma(g|s)\mu(s). \tag{13}$$

4.6 Mass of Searchers

Let μ_i denote the fraction of workers with health status i in the stationary health status distribution. The mass of searchers in health status i can be computed by deducting from μ_i the mass of workers which remain employed by a firm at the time of matching:

$$m_{i'} = \mu_{i'} - \phi\rho \sum_{i \in \{h,u\}} q^H_{ii'} \sum_{s \in \mathcal{S}} \sum_{g \in \mathcal{G}} \sigma_i(s, g)\psi_i(s)\gamma(g|s)\mu(s). \tag{14}$$

4.7 Tightness

By definition tightness is the ratio of the number of vacancies and the mass of searcher, giving rise to the equilibrium relationship

$$\theta = \frac{\phi \sum_{s \in \mathcal{S}} \sum_{g \in \mathcal{G}} g_v \gamma(g|s)\mu(s)}{m}. \tag{15}$$

4.8 Definition of Equilibrium

Definition 1 *$J(\cdot)$, $\gamma(\cdot,\cdot)$, $w(\cdot,\cdot)$, $\mathcal{G}(\cdot)$, $V_h(\cdot)$, $V_u(\cdot)$, $C_h(\cdot,\cdot)$, $C_u(\cdot,\cdot)$, V^s_h, V^s_u, $\mu(\cdot)$, m, ν, ϕ, and θ constitute a stationary equilibrium if they satisfy equations* (2), (6), (7), (8), (9), (10), (11), (12), (14), (15).

An algorithm for computing an equilibrium is outlined in Appendix B.

5 Quantitative Analysis

5.1 Preliminary Calibration

5.1.1 Functional Forms

To conduct quantitative analysis we must choose functional forms for the utility, production, and matching functions and assign parameter values.

Utility Function. We use a constant absolute risk aversion utility (CARA) function[11]

$$U(c_t) = -\frac{\exp(-\varsigma c_t) - 1}{\varsigma} \tag{16}$$

Production Function. Output of a firm with productivity z is given by

$$zF(g_e) = zg_e^{\eta}$$

where g_e is the number of workers. We assume that $\eta \in (0, 1)$ so that the production function exhibits decreasing returns to scale and satisfies the usual Inada conditions.[12] The parameter z varies across establishments and across time generating a cross-sectional and time-series variation in establishment productivity.

Following Hopenhayn and Rogerson (1993), we assume that the productivity shocks evolve according to the process

$$\ln(z') = \zeta(1 - \phi) + \phi \ln(z) + \epsilon', \tag{17}$$

where $0 \geq \phi < 1$, $\zeta \geq 0$ and $\epsilon' \sim N(0, \sigma_\epsilon^2)$. We denote the transition function for z as $Q(z, dz')$. This process has a parsimonious representation with the parameters corresponding to objects that are of intuitive interest given the nature of the employer provided health insurance system that we study. For example, the volatility and persistence of this process have important impact in the variability of insurance provision at

[11] We use CARA instead of the more standard CRRA utility function since it facilitates some steps in the computation.

[12] Our model is a single-good model in which a non-degenerate distribution of establishment sizes is sustained by decreasing returns at the establishment level. An alternative framework is to assume differentiated products and constant returns at the establishment level. In this alternative framework, the nondegenerate distribution of establishment sizes is sustained by curvature in preferences. As discussed in, e.g. Restuccia and Rogerson (2008), conceptually these frameworks are very similar.

the firm level. More complicated process can easily be incorporated into the analysis, but this one appears a reasonable first pass and is a standard process in the firm dynamics literature.

Matching Function. For comparability with much of the literature we choose the Cobb-Douglas functional form of the matching function between workers and firms:

$$M(m, v) = \chi m^{\alpha} v^{1-\alpha}. \tag{18}$$

Thus, the are two parameters, χ, α, that characterize the matching function.

5.1.2 Parameters

The model parameters to be calibrated are:

1. β – the time discount rate,
2. ρ – the probability of a worker leaving the labor market,
3. ς – coefficient of absolute risk aversion,
4. η – curvature of the production function,
5. ζ – unconditional mean of idiosyncratic productivity,
6. ϕ – persistence of idiosyncratic productivity,
7. σ_{ϵ} – st. dev. of innovations in idiosyncratic productivity,
8. χ – scale parameter of the matching function,
9. α – elasticity of the matching function,
10. c_f – firms' flow cost of operating,
11. c – cost of maintaining a vacancy,
12. b – value of non-market activity,
13. τ – proportional tax on labor income,
14. q_I – expiration probability of insurance,
15. $\kappa(g_e)$ – administrative load schedule for employer provided insurance,
16. q^H – the health status transition matrix,
17. e_u, e_h – health expenditures.

We choose the model period to be one month in order to accommodate the high frequency of transitions in the labor market. The data used to compute some of the targets have monthly, quarterly or annual frequency, and we aggregate the model-generated data appropriately when matching those targets.

We choose a relatively low coefficient of absolute risk aversion of $\rho = 0.5$ since the model does not include other channels of insurance beyond health insurance. We choose $\rho = 0.9979$ to generate an expected working lifetime of 40 years. We set $\beta = 1/(1+r)$, where r corresponds to an annual interest rate of 4%.

The extent of decreasing returns in the establishment-level production function is an important parameter in our analysis. As described in Restuccia and Rogerson (2008), direct estimates of establishment-level production functions and different calibration procedures point to a value for $\eta = 0.85$.

Hagedorn and Manovskii (2008) find the average monthly job finding rate of 0.45, and the average value for labor market tightness $\theta = 0.634$. They also estimate that the average flow cost of posting a vacancy equals 58% of the average labor productivity, i.e., $c = 0.58p$. Labor productivity p is defined as output per worker. We target these values.

The two parameters, χ, α, that characterize the matching function are selected to match the average job-finding probability and the elasticity of the job-finding probability with respect to labor market tightness. Petrongolo and Pissarides (2001) survey the empirical evidence and conclude that the value of $\alpha = 0.5$ for the elasticity of the job-finding rate with respect to labor market tightness is appropriate. (See also Brügemann (2008).)

There is only limited evidence on administrative costs (marketing, billing, employee enrollment and education, payments to benefit consultants and insurance sales agents, risk charges, underwriting, etc.) and how they vary with firm size. Congressional Research Service (1988) reports that administrative costs represent 8% of premiums on average, but up to 40% for small firms.[13] Our benchmark model assumes that the ad-

[13] A more recent study by General Accounting Office (GAO) finds that the administrative costs represent 20 to 25% of small employer's premiums compared to 10% for large employers (using 50 employees as the cut-off determining the firm size). These differences may have narrowed due to the growth in managed care and improvements in information technology.

ministrative load is independent of firm size. Thus, we set $\kappa_{g_e} = 0.08$ and $\kappa_1^{ind} = 0.4$ so that the load on individual insurance is similar to that faced by small firms.

In our benchmark calibration we use estimates of the average marginal tax on labor provided by Lucas (1990) and Prescott (2004) and set the proportional tax rate on labor $\tau = 0.4$.[14]

To calibrate the health expenditure process we use data from the Medical Expenditure Panel Survey (MEPS). The MEPS is based on a series of national surveys conducted by the U.S. Agency for Health Care Research and Quality (AHRQ). It consists of two-year panels since 1996/1997 and includes data on demographics, income, health expenditures and insurance. In the model idiosyncratic health shocks follow a two state Markov process: high h or low u with transition probabilities $q^H_{ii'}$ for $i, i' \in \{h, u\}$. In the data we identify health status with medical expenditures $e_u > e_h$. The distribution of health expenditures is very skewed in the data. To approximate it with a two state Markov process we divide workers into two bins. We put the cutoff at the 90th percentile of health expenditures. This choice implies $e_u = .4054$, $e_h = .0274$, where health expenditures are expressed as a function of the mean wage in the economy. The implied transition matrix is given by $q^H_{hh} = 0.9905$, $q^H_{hu} = 0.026$, $q^H_{uh} = 0.0859$.[15] We assume that all the new labor market entrants are healthy so that $q^0_h = 1$ and $q^0_u = 0$.

The mass of operating establishments is pinned down by the free entry condition. All active firms must pay a flow cost c_f in each period of operation. If an establishment does not pay the fixed cost in any period then it ceases to exist. The parameter c_f is important for evaluating the effects of alternative health insurance policies. For example, if $c_f > 0$ potential health insurance policies that subsidize smaller and less efficient firms may induce selection that lowers the aggregate productivity.

At the moment we were only able to compute a model that allows for a maximum firm size of 150 workers, a constraint we plan to relax in the future. Once we are able to compute the model that allows for the existence of large firms we will calibrate the idiosyncratic productivity process to match statistics describing the the size distribution

[14] Mendoza, Razin, and Tesar (1994) report a lower value $\tau = 0.3$ for the average labor tax rate.

[15] The transition matrix in the data is computed at an annual frequency and is given by $q^H_{hh} = 0.93$ and $q^H_{uh} = 0.63$.

of firms based on 1997 US Economic Census and provided in, e.g., Guner, Ventura, and Yi (2008). Since these statistics are not meaningful in the model that only allows for the existence of small firms we choose the parameters of the process in line with the estimates reported in the literature. We chose $\phi = 0.9975$ and $\sigma_\epsilon = 0.02$.

Remaining Parameters. Three parameters remain to be determined: the values of non-market activity, b, the matching function parameter, χ, and the unconditional mean of idiosyncratic productivity shocks process ζ. We choose the values for these parameters to match the data on the average value for labor market tightness, the average values for the job-finding rate, and to generate the normalized value of wages equal to one. Thus, there are three targets, all described in the previous paragraphs, to pin down three parameters. To find the values of these parameters we solve the model numerically according to the computational algorithm described in Appendix II.

Since the average flow cost of posting a vacancy is a function of the equilibrium output per worker, it exact value is pinned down only once the equilibrium size distribution of firms is determined.

5.2 Benchmark Results

[TO BE WRITTEN]

5.3 Evaluating Alternative Policies

[TO BE WRITTEN]

Table 1: Calibrated Parameter Values.

Parameter	Definition	Value
β	the time discount rate	0.996
ς	coefficient of absolute risk aversion	0.5
ρ	the probability of a worker staying in the labor market	0.9979
η	curvature of the production function	0.85
ζ	unconditional mean of idiosyncratic productivity	x.xxx
ϕ	persistence of idiosyncratic productivity	0.9975
σ_ϵ	st. dev. of innovations in idiosyncratic productivity	0.02
χ	scale parameter of the matching function	x.xxx
α	elasticity of the matching function	0.5
c_f	firms' flow cost of operating	0.0
c	cost of maintaining a vacancy	0.58p
b	value of non-market activity	x.xxx
τ	proportional tax on labor income	0.4
κ_{g_e}	administrative load schedule for employer provided insurance	0.08
e_u	health expenditures of unhealthy workers	0.4054
e_h	health expenditures of healthy workers	0.0274
q^H_{hh}	the health status transition probability	0.9905
q^H_{uh}	the health status transition probability	0.0859
q^H_{uu}	the health status transition probability	0.37
q^0_h	fraction of new entrants who are healthy	1.00
q^0_u	fraction of new entrants who are unhealthy	0.00

Note - The table contains the calibrated parameter values in the benchmark calibration.

6 Conclusion

In this paper we attempt to understand the workings of the US health care system. This system is largely employer based. The dominance of employer provided insurance is typically attributed to its tax deductibility, which is not available to workers purchasing insurance individually. Employers are required to provide the same insurance options at the same cost to all employees. Insurance contracts have a typical duration of one year and insurance companies are allowed to change the rates with few binding restrictions.

One prominent feature of the data is that smaller firms are less likely to provide coverage than larger firms. The reasons for this are unclear. This might be due to the existence of fixed cost of obtaining and maintaining coverage. Alternatively, it is possible that since small firms are less able to pool risk among their workers, making it more difficult for them to provide insurance relative to large firms and forcing them to discontinue coverage occasionally.

We develop a quantitative equilibrium model that features tax deductibility of employer-provided coverage, non-discrimination restrictions, fixed costs of coverage and firms that hire discrete numbers of workers in frictional labor markets. We use the calibrated model to understand what drives the patterns of insurance provision with firm size and to evaluate the effects of this system on the flows of workers across health insurance coverage status, labor market flows, as well as on the the size distribution of firms, and aggregate productivity.

This model and the algorithm we developed to compute it represent interesting contributions in themselves and could be applied to study other issues such as, the fragility of small groups to the decisions of key members to join or leave them (the rise and fall of some Economics departments is an example).

As mentioned above, it is not uncommon to receive insurance through ones spouses employer. Thus introducing couples may be an important extension. An interesting question is whether large firms subsidize small firms through this channel.

References

BRÜGEMANN, B. (2008): "What Elasticity of the Matching Function is consistent with U.S. Aggregate Labor Market Data?," Working paper, Yale University.

COCHRANE, J. H. (1995): "Time-Consistent Health Insurance," *Journal of Political Economy*, 103(3), 445–473.

CONGRESSIONAL RESEARCH SERVICE (1988): *Costs and Effects of Extending Health Insurance Coverage*. U.S. Government Printing Office, Washington, DC.

CUTLER, D. M. (1994): "Market Failure in Small Group Health Insurance," Working Paper 4879, National Bureau of Economic Research.

DEY, M. S., AND C. J. FLINN (2005): "An Equilibrium Model of Health Insurance Provision and Wage Determination," *Econometrica*, 73(2), 571–627.

FELDSTEIN, M. S. (1973): "The Welfare Loss of Excess Health Insurance," *Journal of Political Economy*, 81(2), 251–280.

GENERAL ACCOUNTING OFFICE (GAO) (2001): "Private Health Insurance: Small Employers Continue to Face Challenges in Providing Coverage," GAO-02-8.

GUNER, N., G. VENTURA, AND X. YI (2008): "Macroeconomic Implications of Size-Dependent Policies," *Review of Economic Dynamics*, 11(4), 721–744.

HAGEDORN, M., AND I. MANOVSKII (2008): "The Cyclical Behavior of Equilibrium Unemployment and Vacancies Revisited," *American Economic Review*, 98(4), 1692–1706.

HENRY J. KAISER FAMILY FOUNDATION (1998): "The Uninsured: A Primer," Report No. 7451-04, The Kaiser Commission on Medicaid and the Uninsured.

HOPENHAYN, H., AND R. ROGERSON (1993): "Job Turnover and Policy Evaluation: A General Equilibrium Analysis," *Journal of Political Economy*, 101(5), 915–938.

Jeske, K., and S. Kitao (2008): "U.S. Tax Policy and Health Insurance Demand: Can a Regressive Policy Improve Welfare?," *Journal of Monetary Economics*, Forthcoming.

Joint Committee on Taxation (2008): "Tax Expenditures for Health Care," JCX-66-08.

Long, S. H., and M. S. Marquis (1997): *ROBERT WOOD JOHNSON FOUNDATION EMPLOYER HEALTH INSURANCE SURVEY [COMMUNITY TRACKING STUDY AND STATE INITIATIVES IN HEALTH CARE REFORM PROGRAM], 1997 [Computer file].* ICPSR02935-v2. Washington, DC: RAND Corporation [producer], 2000. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2005-07-06. doi:10.3886/ICPSR02935.

——— (1998): "Trends in Offering Employer-Sponsored Coverage," Data Bulletin No. 15, Center for Studying Health Care Change.

Lucas, R. J. (1990): "Supply Side Economics: An Analytical Review," *Oxford Economic Papers*, 42(2), 293–316.

Lyke, R. (2008): "The Tax Exclusion for Employer-Provided Health Insurance: Policy Issues Regarding the Repeal Debate," *Congressional Research Service Report for Congress*, RL34767.

Mendoza, E., A. Razin, and L. Tesar (1994): "Effective Tax Rates in Macroeconomics: Cross-country Estimates of Tax Rates on Factor Incomes and Consumption," *Journal of Monetary Economics*, 34, 297–323.

Office of Management and Budget (2008): *Analytical Perspectives, Budget of the United States Government, Fiscal Year 2009*. U.S. Government Printing Office, Washington, DC.

Petrongolo, B., and C. Pissarides (2001): "Looking into the Black Box: A Survey of the Matching Function," *Journal of Economic Literature*, 39(2), 390–431.

Prescott, E. (2004): "Why Do Americans Work So Much More Than Europeans," *Federal Reserve Bank of Minneapolis Quarterly Review*, 28(1), 2–13.

Restuccia, D., and R. Rogerson (2008): "Policy Distortions and Aggregate Productivity with Heterogeneous Plants," *Review of Economic Dynamics*, 11(4), 707–720.

Selden, T. M., and B. M. Gray (2006): "Tax Subsidies for Employer-Related Health Insurance: Estimates for 2006," *Health Affairs*, 25(2), 1568–1579.

Thomasson, M. A. (2002): "From Sickness to Health: The Twentieth Century Development of U.S. Health Insurance," *Explorations in Economic History*, 39(3), 233–253.

——— (2003): "The Importance of Group Coverage: How Tax Policy Shaped U.S. Health Insurance," *American Economic Review*, 93(4), 1373–1384.

APPENDICES

I Transition Probabilities

I.1 Firm Transition $\mu(s'|s, g, \theta, \nu)$

Consider a firm in state s with policy g. After workers have been induced to leave (g_h, g_u) workers remain, so at this stage the workforce of the firm is still deterministic. Next workers are dismissed at random, leaving g_e workers in total. The probability of arriving at the workforce (ψ_h, ψ_u) after random dismissal is given by

$$q^{dis}(\psi_h, \psi_u|g) = \begin{cases} \frac{\binom{g_h}{\psi_h}\binom{g_u}{\psi_u}}{\binom{g_h+g_u}{g_e}}, & \text{if } \psi_h + \psi_u = g_e, \\ 0, & \text{otherwise.} \end{cases} \tag{A1}$$

The logic behind this formula is as follows. The total number of workers before random dismissal is $g_h + g_u$, and there are g_e slots. There are $\binom{g_h+g_u}{g_e}$ different ways of allocating these slots, all equally likely. The number of different ways of allocating these slots which have ψ_h healthy and ψ_u unhealthy workers are $\binom{g_h}{\psi_h}\binom{g_u}{\psi_u}$. Vandermonde's identity insures that these probabilities add up to one. Let $Q^{dis}(g)$ collect these probabilities in a vector, ordering workforces in the natural way: $(0,0)$, $(1,0)$, $(0,1)$, $(2,0)$, $(1,1)$, $(0,2)$ and so on.

Next workers draw their new health status, and we need to compute the probability of transiting from (ψ_h, ψ_u) to (ψ'_h, ψ'_u) in this step. The number of workers must remain the same $\psi'_h + \psi'_u = \psi_h + \psi_u$, otherwise the probability of this transition is zero. For a transition to (ψ'_h, ψ'_u) it must be that the number of workers remaining in status h is at least $\max\{\psi'_h - \psi_u, 0\}$, because no more than ψ_u can join from status u. The probability of j workers remaining in status h is given by $B\left(j; \psi_h, q^H_{hh}\right)$. Here B is the binomial distribution: the first argument is the number of successes, the second argument the number of trials, and the third argument the probability of success. If j workers remain in status h, the transition to ψ'_h requires that exactly $\psi'_h - j$ switch from status u to status h. The latter happens with probability $B\left(\psi'_h - j; \psi_u, q^H_{uh}\right)$. Thus

$$q^{health}(\psi_h, \psi_u; \psi'_h, \psi'_u) = \sum_{j=\max\{\psi'_h - \psi_u, 0\}}^{\min\{\psi'_h, \psi_h\}} B\left(j; \psi_h, q^H_{hh}\right) B\left(\psi'_h - j; \psi_u, q^H_{uh}\right).$$

Notice that this probability does not depend on the firm's policy g. Let Q^{health} denote the transition matrix associated with this step.

In the next step workers exit the labor market with probability $1-\rho$, or quit exogenously with probability δ. Thus a worker stays with the firm with probability $\rho(1-\delta)$. The probability of a transition from (ψ_h, ψ_u) to (ψ'_h, ψ'_u) in this step is

$$q^{exit}\left(\psi_h, \psi_u; \psi'_h, \psi'_u\right) = B(\psi'_h; \psi_h, \rho(1-\delta))B(\psi'_u; \psi_u, \rho(1-\delta)).$$

Let Q^{exit} denote the transition matrix.

The final step is that searching workers are allocated to the firm. The firm has g_v vacancies, each of which is filled with probability $q(\theta)$. The probability to transit from (ψ_h, ψ_u) to (ψ'_h, ψ'_u) is

$$\begin{aligned} & q^{vac}\left(\psi_h, \psi_u; \psi'_h, \psi'_u\right) \\ & \equiv B\left(\psi'_h + \psi'_u - \psi_h - \psi_u; g_v, q(\theta)\right) B\left(\psi'_h - \psi_h; \psi'_h + \psi'_u - \psi_h - \psi_u, \nu\right) \end{aligned}$$

The first term captures that out of g_v vacancies it must be that $\psi'_h + \psi'_u - \psi_h - \psi_u$ make contact with a worker, with a probability of success $q(\theta)$. The second term captures that out of these contacts, $\psi'_h - \psi_h$ must be with a healthy worker, with a probability of success ν. Let $Q^{vac}(g, \theta, \nu)$ denote the transition matrix.

Combining these transitions, the distribution $\mu(s'|s, g)$ is given by

$$\mu(s'|s, g) = q^Z_{z(s)z(s')} q^I_{I(s)I(s')}(g) \left[Q^{vac}(g, \theta, \nu) \cdot Q^{exit} \cdot Q^{health} \cdot Q^{dis}(g)\right]_{\psi_h(s')\psi_u(s')}$$

where $[Q]_{\psi_h\psi_u}$ extracts the element of vector Q corresponding to the workforce $\psi_h\psi_u$. The transition probabilities for insurance provision status are $q^I_{CC}(g) = 1 - q^I$, $q^I_{CE}(g) = q^I$, $q^I_{EC}(g) = g^I$, $q^I_{EN}(g) = 1 - g^I$, $q^I_{NC}(g) = g_I$, $q^I_{NN}(g) = 1 - g_I$, and zero for the remaining transitions.

I.2 Worker Transition $\mu_{ii'}\left[s'|s, g, \theta, \nu\right]$

We derive $\mu_{hh'}\left[s'|s, g, \theta, \nu\right]$, the remaining cases are analogous. The calculations parallel the derivation of the firm transition, with the twist that we need to condition on the worker being healthy both in this period and in the next period, and that the worker remains in the labor market and stays with the firm.

After workers have been induced to leave (g_h, g_u) workers remain. Next workers are dismissed at random, leaving g_e workers in total. Conditioning on the worker staying and being healthy, the probability of arriving at the workforce $\left(\hat{\psi}_h, \hat{\psi}_u\right)$ after random dismissal is given by

$$q_{hh}^{dis}\left(\hat{\psi}_h, \hat{\psi}_u | g\right) = \begin{cases} \frac{\binom{g_h - 1}{\hat{\psi}_h - 1}\binom{g_u}{\hat{\psi}_u}}{\binom{g_h + g_u - 1}{g_e - 1}}, & \text{if } \hat{\psi}_h + \hat{\psi}_u = g_e, \\ 0, & \text{otherwise.} \end{cases} \tag{A2}$$

The logic behind this formula is as follows. The worker under consideration is healthy and is not dismissed. The remaining number of workers at risk of random dismissal is $g_h + g_u - 1$, and there are $g_e - 1$ remaining slots. There are $\binom{g_h+g_u-1}{g_e-1}$ different ways of allocating these slots, all equally likely. To end up at $(\hat{\psi}_h, \hat{\psi}_u)$ it must be that $\hat{\psi}_h - 1$ of these slots go to healthy workers. The number of different ways of allocating these slots which have $\hat{\psi}_h - 1$ healthy and $\hat{\psi}_u$ unhealthy workers are $\binom{g_h-1}{\hat{\psi}_h-1}\binom{g_u}{\hat{\psi}_u}$. Let $Q_{hh}^{dis}(g)$ denote the vector of these probabilities.

Next workers draw their new health status. We compute the probability of a transition from (ψ_h, ψ_u) to (ψ'_h, ψ'_u) The number of workers must remain the same $\psi_h + \psi_u = \psi'_h + \psi'_u$, otherwise the probability of this transition is zero. Here we need to condition on the event that the worker under consideration stays healthy. For a transition to (ψ'_h, ψ'_u) it must be that the number of workers remaining in status h is at least $\max\{\psi'_h - \psi_u, 1\}$, because no more than ψ_u can join from status u, and we already condition on one healthy worker staying healthy. The probability of j workers remaining in status h is given by $B\left(j-1; \psi_h - 1, q_{hh}^H\right)$. If j workers remain in status h, the transition to ψ'_h requires that exactly $\psi'_h - j$ switch from status u to status h. The latter happens with probability $B\left(\psi'_h - j; \psi_u, q_{uh}^H\right)$. Thus

$$q_{hh}^{health}\left(\psi_h, \psi_u; \psi'_h, \psi'_u\right) = \sum_{j=\max\{\tilde{\psi}_h - \hat{\psi}_u, 1\}}^{\min\{\tilde{\psi}_h, \hat{\psi}_h\}} B\left(j-1; \hat{\psi}_h - 1, q_{hh}^H\right) B\left(\tilde{\psi}_h - j; \hat{\psi}_u, q_{uh}^H\right).$$

Let Q_{hh}^{health} denote the associated transition matrix.

Next workers exit the labor force with probability $(1-\rho)$, or separate exogenously with probability δ. We condition on the worker under consideration remaining with the firm. The probability of a transition from (ψ_h, ψ_u) to (ψ'_h, ψ'_u) in this step is

$$q_{hh}^{exit}(\psi_h, \psi_u; \psi'_h, \psi'_u) = B(\psi'_h - 1; \psi_h - 1, \rho)B(\psi'_u; \psi_u, \rho).$$

Let Q_{hh}^{exit} denote the transition matrix.

Again, the final step is that searching workers are allocated to the firm. This step is not affected by conditioning.

Combining these transitions, the distribution $\mu(s'|s, g, \theta, \nu)$ is given by

$$\mu(s'|s, g, \theta, \nu) = q^Z_{z(s)z(s')} q^I_{I(s)I(s')}(g) \left[Q^{vac}(g, \theta, \nu) \cdot Q_{hh}^{exit} \cdot Q_{hh}^{health} \cdot Q_{hh}^{dis}(g)\right]_{\psi_h(s'), \psi_u(s')}$$

I.3 Searching Worker Transition $\mu^s_{i'}[s'|\mu(\cdot), \gamma(\cdot|\cdot), \theta, \nu]$

A searching worker who makes contact is randomly allocated to a vacancy. Let s denote the state of the worker's new firm at the beginning of the period when it posted the vacancy. This firm implements policy g with probability $\gamma(g|s)$, in which case it has g_v vacancies. Thus the searcher is matched with a firm in state s implementing policy g with probability

$$q^{match}[s, g|\mu(\cdot), \gamma(\cdot|\cdot)] = \frac{g_v \gamma(g|s)\mu(s)}{\sum_{\tilde{g}\in G}\sum_{s\in\mathcal{S}} \tilde{g}_v \gamma(\tilde{g}|s)\mu(s)}.$$

If matched with a firm implementing policy g, the distribution of that firm's workforce after random dismissal and exogenous separations is $Q^{dis}(g)$, and after health status changes and exit from the labor market it is $Q^{exit} \cdot Q^{health} Q^{dis}(g)$. Next the firm is allocated searchers. In this step we need to condition on the worker under consideration having new health status i'. Here we consider the case $i' = h$, the case $i' = u$ is analogous. If a firm has workforce (ψ_h, ψ_u) before vacancies are filled and follows policy g, then the probability to arrive at (ψ'_h, ψ'_u) is

$$\begin{aligned} q_h^{vac}(\psi_h, \psi_u; \psi'_h, \psi'_u|g) \equiv\ & B\left(\psi'_h + \psi'_u - \tilde{\psi}_h - \tilde{\psi}_u - 1; g_v - 1, q(\theta)\right) \\ & \cdot B\left(\psi'_h - \tilde{\psi}_h - 1; \psi'_h + \psi'_u - \tilde{\psi}_h - \tilde{\psi}_u - 1, \nu\right). \end{aligned}$$

The worker under consideration has already filled one vacancy and is healthy. The first term captures that out of $g_v - 1$ remaining vacancies it must be that $\psi'_h + \psi'_u - \tilde{\psi}_h - \tilde{\psi}_u - 1$

make contact with a worker, with a probability of success $q(\theta)$. The second term captures that out of these contacts, $\psi_h' - \tilde{\psi}_h - 1$ must be with a healthy worker, with a probability of success ν. Let $Q_h^{vac}(g, \theta, \nu)$ denote the associated transition matrix. Then the distribution after this step is $Q_h^{vac}(g, \theta, \nu) \cdot Q^{exit} \cdot Q^{health} Q^{dis}(g)$ Combining these three steps

$$\mu_h^s [s' | \mu(\cdot), \gamma(\cdot|\cdot), \theta, \nu] = \sum_{s \in \mathcal{S}} \sum_{g \in \mathcal{G}} \Big\{ \left[Q_h^{vac}(g, \theta, \nu) \cdot Q^{exit} \cdot Q^{health} \cdot Q^{dis}(g) \right]_{\psi_h(s')\psi_u(s')} \cdot q^Z_{z(s)z(s')} q^I_{I(s)I(s')}(g) q^{match} [s, g | \mu(\cdot), \gamma(\cdot|\cdot)] \Big\}$$

II Computational Algorithm

The algorithm iterates on the policy function $\gamma(\cdot|\cdot)$ until convergence.

First, notice that for any policy function it is straightforward to compute $\mathcal{J}(\cdot)$, $V_h(\cdot)$, $V_u(\cdot)$, V_h^s, V_u^s, $\mu(\cdot)$, m, ν, and θ consistent with that policy. Second, for any $\{\mathcal{J}(\cdot), V_h(\cdot), V_u(\cdot), V_h^s, V_u^s, \mu(\cdot), m_h, m_u, \theta\}$ we can use equation (8) to compute a set of policies which are optimal. Combining these two mappings, we get a correspondence Ω mapping policy functions into sets of policy functions. Stationary equilibrium policy functions are the fixed points of this correspondence, so we're looking for $\gamma(\cdot|\cdot)$ such that

$$\gamma(\cdot|\cdot) \in \Omega[\gamma(\cdot|\cdot)].$$

For a policy function $\gamma^k(\cdot|\cdot)$, let $\gamma^k(\cdot|\cdot; \tilde{\gamma}(\cdot|s))$ denote the policy given by

$$\gamma^k(\cdot|z', \psi_h', \psi_u'; \tilde{\gamma}(\cdot|s)) = \begin{cases} \gamma^k(\cdot|z', \psi_h', \psi_u') & \text{for all} \quad (z', \psi_h', \psi_u') \neq (s), \\ \tilde{\gamma}(\cdot|s) & \text{for} \quad (z', \psi_h', \psi_u') = (s). \end{cases}$$

In words, $\gamma^k(\cdot|\cdot; \tilde{\gamma}(\cdot|s))$ is obtained from $\gamma^k(\cdot|\cdot)$ by switching out the policy at one point in the state space, replacing $\gamma^k(\cdot|s)$ with $\tilde{\gamma}(\cdot|s)$.

Given $\Omega[\gamma(\cdot|\cdot)]$, define the projection

$$\Omega_{(s)}[\gamma(\cdot|\cdot)] = \{\tilde{\gamma}(\cdot|s) \in \Gamma | \tilde{\gamma}(\cdot, \cdot) \in \Omega[\gamma(\cdot|\cdot)]\}.$$

This is the sets of mixed policies that are optimal for the point in the state space (s) given that all other equilibrium objects are induced by the policy function $\gamma(\cdot|\cdot)$.

The algorithm starts with a guess $\gamma^0(\cdot|\cdot)$. The approach is to find a fixed point for just one point in the state space at each iteration, and to move randomly through the state space until convergence. Iteration k comprises the following steps:

1. Pick a point in the state space $(z^k, \psi^k) \in \mathcal{S}$ at random.

2. Given the policy function $\gamma^k(\cdot|\cdot)$, find a mixed policy $\tilde{\gamma}(\cdot|z^k, \psi^k)$ such that

$$\gamma^k(\cdot|\cdot; \tilde{\gamma}(\cdot|z^k, \psi^k)) \in \Omega_{(z^k,\psi^k)}[\gamma^k(\cdot|\cdot; \tilde{\gamma}(\cdot|z^k, \psi^k))].$$

   This step is implemented using a heuristic algorithm.

3. Set $\gamma^{k+1}(\cdot, \cdot) = \gamma^k(\cdot|\cdot; \tilde{\gamma}(\cdot|z^k, \psi^k))$.