Fragility: A Quantitative Analysis of the US Health Insurance System

Björn Brügemann† Iourii Manovskii‡
Yale University University of Pennsylvania

First version: April 28, 2008
This version: February 15, 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.


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. 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 42.6 percent of establishments in firms of fewer than 50 workers offer health insurance to any of their workers, compared with 95.6 percent of establishments in firms of 50 or more workers (Medical Expenditure Panel Survey (2006)). 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.” Consistent with this theory, Cutler (1994) finds that small employers experience greater year-to-year variabili-

---

1Under 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)).
ity in medical expenses than do large firms. The relative importance of these two (and possibly other) channels in explaining the observed variation of insurance provision with firm size is not 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. 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. 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.

---

2The 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. Thomasson (2002, 2003) and Lyke (2008) describe the historical evolution of the US health insurance system.
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 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)). 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

---

4Interestingly, over 80 percent of the uninsured are wage earners or members of working families.

5Long and Marquis (1998) document, using the large nationally representative 1997 Robert Wood Johnson Foundation Employer Health Insurance Survey, that a sizable number of employers drop their health coverage. For example, almost one-quarter of the firms with less than 10 employees that offered insurance in 1995 did not offer it in 1997, as compared to only 4 percent of firms with more than 100 employees. Similarly, 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.
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 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 develop the model of the employer-provided health insurance coverage that features the fragility channel. In Section 3 we define equilibrium. In Section 4 we calibrate the model and perform the quantitative analysis of the current system and evaluate the effects of several proposed reforms. Section 5 concludes.

2 Model

There is mass \( \phi \) of firms. 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.

Preferences of workers are

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

with \( U' > 0 \) and \( U'' \leq 0 \). Firms share the discount factor \( \beta \) 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 \( q^{zz'} \).

Output of a firm with productivity \( z \) is given by

\[
z F(g_e)
\]

where \( g_e \) is the number of workers and \( F' > 0, F'' \leq 0 \).

Workers are subject to idiosyncratic shocks to their health. Their health status follows a two state Markov process: high (healthy) \( h \) or low (unhealthy) \( l \) with transition probabilities \( q^{ii'}_i \) for \( i, i' \in \{h, l\} \). We assume persistence \( q^{hh}_h > q^{lh}_l \). Health status in turn determines medical expenditures \( e_l > e_h \). 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 health 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.
A firm can reduce its workforce in two ways. First, if workers of a given health status are indifferent about staying with the firm, then the firm is free to choose the number of workers of that type that remain. Second, if workers of a given health status strictly prefer to stay with the firm, the firm can still attempt to reduce the number of workers of that type, but it has to do so in a non-discriminatory fashion. Specifically, it has to choose workers to be dismissed at random.

A firm can recruit new workers by posting vacancies. It posts $g_v \in \{0, 1, \ldots \}$ vacancies at cost $c$ 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

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

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

Health insurance contracts offered to firms last for one period, longer term contracts are not available. Premiums reflect health expenditures on actuarially fair terms plus an administrative load which is a function of firm size.

Analogous contracts are available for workers to purchase individually, with administrative costs that differ from group insurance. Individuals, however, have to pay premiums out of after tax wages. The tax rate on labor income is denoted $\tau \in (0, 1)$.

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

## 3 Stationary Equilibrium

The state of a firm at the beginning of the period $(z, \psi)$ is given by its productivity $z \in \mathcal{Z}$ and workforce $\psi = (\psi_h, \psi_l) \in \Psi$. Let $\mathcal{S} = \mathcal{Z} \times \Psi$ denote the state space.
A stationary equilibrium consists of the following objects:

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

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

3. Worker value functions $V_h(\cdot), V_l(\cdot) : S \to \mathbb{R}$, with $V_i(z, \psi)$ giving the utility of a worker with health status $i \in \{h, l\}$ employed in a firm in state $(z, \psi)$.

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

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

6. Masses of workers searching $m_h, m_l \in [0, 1]$, in good and poor health, respectively.

7. 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.

### 3.1 Firm Decision

Consider a firm in state $(z, \psi) \in S$. Firms do not benefit from paying workers more than their opportunity cost. The constraint of a common wage, however, implies that in general a firm can only make workers in one health status indifferent. This reduces the wage decision to a binary choice of the health status left indifferent $g_w \in \{h, l\}$. A second binary choice is whether to provide insurance $g_I \in \{0, 1\}$. The firm can induce any number of workers with a given health status to leave as long as workers in this status weakly prefer to leave. Let $g_h$ and $g_l$ denote the numbers of healthy and unhealthy workers left after the firm has exercised this option, respectively. The firm can also dismiss workers at random. Let $g_e \in \{0, 1, \ldots, g_h + g_l\}$ denote the total number of workers left after random dismissal. Finally, the firm has to decide on the number of vacancies to
post $g_v \in \{0, 1, \ldots\}$. Let $g = (g_w, g_I, g_h, g_t, g_e, g_v)$ denote a pure policy of the firm. Let $\mathcal{G}$ denote the set of all pure policies which are consistent with the upper bound on firm size

$$\mathcal{G} \equiv \{(g_w, g_I, g_h, g_t, g_e, g_v) | g_w \in \{h, l\} \land g_I \in \{0, 1\} \land (g_h, g_t, g_e, g_v) \in \mathbb{N}_0^4 \land g_e + g_v \leq \bar{\psi}\}.$$ 

Define the set of all mixed policies

$$\Gamma = \left\{ \gamma(\cdot) : \mathcal{G} \rightarrow [0, 1] \left| \sum_{g \in \mathcal{G}} \gamma(g) = 1 \right. \right\}.$$ 

Let $\mathcal{G}(z, \psi) \subset \mathcal{G}$ denote the set of pure policies which are feasible for a firm in state $(z, \psi)$. This set will be specified precisely below. Given this set, the set of mixed policies feasible for a firm in this state, denoted $\Gamma(z, \psi) \subset \Gamma$, is given by

$$\Gamma(z, \psi) \equiv \left\{ \gamma(\cdot) : \mathcal{G} \rightarrow [0, 1] \left| \sum_{g \in \mathcal{G}(z, \psi)} \gamma(g) = 1 \land \sum_{g \in \mathcal{G} \setminus \mathcal{G}(z, \psi)} \gamma(g) = 0 \right. \right\}.$$ 

For a pure policy $g \in \mathcal{G}$ let $V_i(z, \psi|g)$ denote the utility of workers with health status $i$ at the stage within the period just before random dismissal. The flow utility of a worker that remains with the firm at the time of production is

$$u_i(z, \psi|g) = g_I U((1 - \tau)w(z, \psi|g)) + (1 - g_I) \max \left\{ U((1 - \tau)w(z, \psi|g) - p_i), q_{ih}^H U((1 - \tau)w(z, \psi|g) - e_h) + q_{il}^H U((1 - \tau)w(z, \psi|g) - e_l) \right\}.$$ 

The first term is the flow utility of the worker if insurance is provided by the employer. The second term gives the maximum flow utility from two options, individual health insurance and going uninsured. Here

$$p_i = (1 + \kappa_{1\text{ind}})(q_{ih}^H e_h + q_{il}^H e_l)$$

denotes the premium of individual health insurance where $\kappa_{1\text{ind}} \geq 0$ is the administrative load. Similarly, the premium of employer provided insurance, denoted $p(z, \psi|g)$, is given by

$$p(z, \psi|g) \equiv (1 + \kappa_{ge}) \frac{g_h(q_{hh}^H e_h + q_{hl}^H e_l) + g_t(q_{lh}^H e_h + q_{ll}^H e_l)}{g_h + g_t}.$$
A pure policy \(g\) induces a distribution over the firm’s future workforce denoted \(\mu_{ii'}(\psi'|g)\) 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. The probability that a worker remains after random dismissal is

\[
\sigma(g) \equiv \frac{ge}{gh + gi}.
\]

Combining these elements, the utility of a worker with health status \(i\) from staying before random dismissal is

\[
V_i(z, \psi|g) = \sigma(g) \left\{ u_i(z, \psi|g) + \beta \rho \left[ q_{ih} \sum V_{ih}(z', \psi') q_{zz} \mu_{ih}(\psi'|g) + q_{il} \sum V_{il}(z', \psi') q_{zz} \mu_{il}(\psi'|g) \right] \right\} + (1 - \sigma(g)) V_s^i.
\]

(2)

For the pure policy \(g\) to be feasible it must be that workers not induced to leave weakly prefer to stay:

\[
V_i(z, \psi|g) \geq V_s^i \quad \text{if} \quad g_i > 0.
\]

(3)

However, if \(g_i < \psi_i\) 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. For \(g\) to be feasible it must be that

\[
V_i(z, \psi|G^+_i(g)) \leq V_s^i \quad \text{if} \quad g_i < \psi_i.
\]

(4)

Thus the set of feasible policies for a firm in state \((z, \psi)\) is

\[
\mathcal{G}(z, \psi) = \{ g \in \mathcal{G}| gh \leq \psi_h \wedge gi \leq \psi_l \wedge (3), (4) \text{ hold} \}.
\]

(5)

By definition of the policy \(g \in \mathcal{G}\) it must be that workers in status \(g_w\) are indifferent between staying and leaving, that is \(V_{g_w}(z, \psi|g) = V_{g_w}^s\). Through this restriction equation (2) implicitly defines the wage \(w(z, \psi|g)\). Using this wage one obtains the utility of workers in the other health status.
Flow profits of a firm are given by output minus wage payments minus health insurance premiums minus recruiting costs:

\[ \pi(z, \psi|g) = zF(g_e) - w(z, \psi|g)g_e - gIp(z, \psi|g)g_e - cg. \]

A pure policy \( g \) induces a distribution over the firm’s future workforce \( \mu(\psi'|g) \), which is derived in Appendix I.1. The firm value function \( J(\cdot) \) must satisfy the Bellman equation

\[ J(z, \psi) = \max_{g \in G(z, \psi)} \left\{ \pi(z, \psi|g) + \beta \sum_{(z', \psi') \in S} J(z', \psi')q_{zz'}\mu(\psi'|g) \right\} \quad \text{for all } (z, \psi) \in S. \tag{6} \]

The policy function \( \gamma(\cdot) \) must satisfy

\[ \gamma(g|z, \psi) > 0 \Rightarrow g \in \arg \max_{g \in G(z, \psi)} \left\{ \pi(z, \psi|g) + \beta \sum_{(z', \psi') \in S} J(z', \psi')q_{zz'}\mu(\psi'|g) \right\} \tag{7} \]

for all \( g \in G \) and all \((z, \psi) \in S\).

### 3.2 Worker Value Functions

The utility of a worker in health status \( i \) induced by the optimal policy must satisfy

\[ V_i(z, \psi) = \sum_{g \in G} \gamma(g|\psi, z) \left[ \left( 1 - \frac{g_i}{\psi_i} \right) V_i^s + \frac{g_i}{\psi_i} \max \{ V_i(z, \psi|g), V_i^s \} \right]. \tag{8} \]

### 3.3 Value of Searching

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

\[ u_i^s = \max \{ U(b - p_i), q_i^H U(b - e_h) + q_i^H U(b - e_l) \}, \]

which is the better option of individual insurance and going uninsured, and \( b \) denotes the value of non-market activity.

Utility of this worker is

\[ V_i^s = u_i^s + \beta \rho f(\theta) \left[ q_i^H \sum V_h(z', \psi') \mu_h^s(z', \psi') + q_i^H \sum V_l(z', \psi') \mu_l^s(z', \psi') \right] + \beta \rho (1 - f(\theta)) \left[ q_i^H V_h^s + q_i^H V_l^s \right] \tag{9} \]
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 $\mu^i(z', \psi')$ 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.

3.4 Invariant Distribution of Firms

The invariant distribution $\mu(z, \psi)$ must satisfy

$$
\mu(z', \psi') = \sum_{(z,\psi)\in S} \sum_{g\in G(z,\psi)} q_{zz'} \mu(\psi'|g) \gamma(g|z,\psi) \mu(z,\psi). \tag{10}
$$

3.5 Mass of Searchers

The timing is such that workers that separate within a period search within that same period. The mass of workers in health status $i$ that separate is given by

$$
\sum_{(z,\psi)\in S} \sum_{g\in G} \left[ \psi_i - g_i(z,\psi) + (1 - \sigma(z,\psi)) g_i(z,\psi) \right] \gamma(g|z,\psi) \mu(z,\psi).
$$

In addition, a mass $(1 - f(\theta)) m_i$ of workers in health status $i$ searched last period but was unsuccessful. Finally, $(1 - \rho) q_0^i$ workers enter the labor market with health status $i$. Health status change and exit from the labor market occurs before matching, thus

$$
m_{i'} = \sum_{i\in\{h,l\}} \rho q_{i' i} \left\{ \sum_{(z,\psi)\in S} \sum_{g\in G} \left[ \psi_i - g_i(z,\psi) + (1 - \sigma(z,\psi)) g_i(z,\psi) \right] \gamma(g|z,\psi) \mu(z,\psi) \right. 
$$

$$
\left. + (1 - f(\theta)) m_i + (1 - \rho) q_0^i \right\} \tag{11}
$$

for $i' \in \{h, l\}$.

3.6 Tightness

The equilibrium mass of vacancies is given by

$$
v = \phi \sum_{(z,\psi)\in S} \sum_{g\in G} g \gamma(g|z,\psi) \mu(z,\psi). \tag{12}
$$

Thus equilibrium tightness must satisfy

$$
\theta = \frac{v}{m_h + m_l}. \tag{13}
$$
3.7 Definition of Equilibrium

\(J(\cdot), \gamma(\cdot, \cdot), V_h(\cdot), V_l(\cdot), \mu(\cdot), V_h^s, V_l^s, m_h, m_l\), and \(\theta\) constitute a stationary equilibrium if

1. \(J(\cdot)\) satisfies equation (6) given \(V_h(\cdot), V_l(\cdot), V_h^s, V_l^s\), and \(\theta\).

2. \(\gamma(\cdot|\cdot)\) satisfies equation (7) given \(V_h(\cdot), V_l(\cdot), V_h^s, V_l^s\), and \(\theta\).

3. \(V_h(\cdot)\) and \(V_l(\cdot)\) satisfy equation (8) given \(\gamma(\cdot|\cdot), V_h^s, V_l^s\).

4. \(V_h^s\) and \(V_l^s\) satisfy equation (9) given \(\gamma(\cdot|\cdot), V_h(\cdot), V_l(\cdot), \mu(\cdot), \theta\).

5. \(\mu(\cdot)\) satisfies equation (10) given \(\gamma(\cdot|\cdot)\) and \(\theta\).

6. \(m_h, m_l\) satisfy equation (11) given \(\gamma(\cdot|\cdot)\) and \(\theta\).

7. \(\theta\) satisfies equation (13) given \(\gamma(\cdot|\cdot), \mu(\cdot), m_h, m_l\).

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

3.8 Conjectures

We conjecture that in equilibrium active continuing firms will be in one of four regimes.

1. Healthy workers are indifferent. Health insurance is offered. Unhealthy workers receive a rent. The labor force is adjusted through random firing. This is better than inducing only healthy workers to leave, since unhealthy workers are more costly. This regime is likely to be adopted if the fraction of unhealthy workers is small. We expect large firms to be in this situation.

2. Healthy workers are indifferent. Health insurance is not offered. All unhealthy workers leave. Additional labor force adjustment is accomplished through inducing health workers to leave or random firing, which are equivalent in this situation. This regime is likely to be adopted for intermediate fractions of unhealthy workers. Given that the composition of their work force is more volatile, we expect small firms to be in this situation more frequently.
3. Unhealthy workers are indifferent. Health insurance is not offered. Healthy workers receive rents. Some unhealthy workers are induced to leave. This is similar to the previous regime, only that it is not optimal to get rid of all the unhealthy workers due to decreasing returns. It may well be that this regime does not occur in equilibrium if decreasing returns are sufficiently weak.

4. Unhealthy workers are indifferent. Health insurance is offered. All healthy workers leave. Here the firm specializes in unhealthy workers, offering them insurance but paying them low wages. The fact that wages are low induces healthy workers to leave. This regime is expected to arise only if the fraction of unhealthy workers is very large.

The third regime is the only one in which healthy workers receive rents. If it does not arise in equilibrium then matters simplify considerably, as the utility of healthy workers always equals the value of search.

\[ V_h(z, \psi) = V_h^s \]

for all \((z, \psi)\).

4 Quantitative Analysis

4.1 Preliminary Calibration

4.1.1 Functional Forms

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

Utility Function. For now we work with a linear specification of utility:

\[ U(c_t) = c_t. \]  \hspace{1cm} (14)

The model and computational programs are set up to allow for curvature in utility but we do not do so at the first pass. Of course, with linear utility workers would not be interested in purchasing group health insurance had it not been subsidized because of tax deductibility.
Production Function. Output of a firm with productivity $z$ is given by

$$zF(g_e) = zg_e^n$$

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. 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

$$\ln(z') = \zeta(1 - \phi) + \phi \ln(z) + \epsilon'$$

where $0 \geq \phi < 1$, $\zeta \geq 0$ and $\epsilon' \sim N(0, \sigma^2)$. 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 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:

$$M(m^h + m^l, v) = \chi(m^h + m^l)^\alpha v^{1-\alpha}.$$  

Thus, the are two parameters, $\chi$, $\alpha$, that characterize the matching function.

4.1.2 Parameters

The model parameters to be calibrated are:

\footnote{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.}
1. $\beta$ – the time discount rate,
2. $\rho$ – the probability of a worker leaving the labor market,
3. $\eta$ – curvature of the production function,
4. $\zeta$ – unconditional mean of idiosyncratic productivity,
5. $\phi$ – persistence of idiosyncratic productivity,
6. $\sigma_\epsilon$ – st. dev. of innovations in idiosyncratic productivity,
7. $\chi$ – scale parameter of the matching function,
8. $\alpha$ – elasticity of the matching function,
9. $c_f$ – firms’ flow cost of operating,
10. $c$ – cost of maintaining a vacancy,
11. $b$ – value of non-market activity,
12. $\tau$ – proportional tax on labor income,
13. $\kappa_{ge}$ – administrative load schedule for employer provided insurance,
14. $\kappa_{ind}^1$ – administrative load for individual insurance,
15. $q^H$ – the health status transition matrix,

We choose the model period to be one quarter. We see it as a compromise between two competing objectives. On the one hand, the lengths of the model period determines the extent of labor market frictions suggesting that the model period should be fairly short, somewhere from one week to one month. On the other hand, the model period also represent the length of time over which insurance contracts are signed. This period typically corresponds to one year in the data. In the near future we will modify the model to uncouple these two objectives (by making the insurance coverage a state variable that gets reset stochastically with the expected duration between resets of one year). For now, however, setting the model period to one quarter appears a reasonable compromise. 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 $\rho = 0.99375$ 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% (this is appropriate in the model with risk-neutral agents).
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 and replace the target quarterly job-finding rate with $\approx 1.0$.

The two parameters, $\chi$, $\alpha$, 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. Our benchmark model assumes that the administrative load is independent of firm size. Thus, we set $\kappa_{ge} = 0.08$ and $\kappa_{1ind} = 0.4$ so that the load on individual insurance is similar to that faced by small firms. Of course, given our current assumption of risk-neutrality, workers would not be interested in purchasing individual health insurance for any $\kappa_{1ind} > 0.0$. The exact value of the load will become relevant when we study the economy with risk-averse workers.

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.

---

7A 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.
\[ \tau = 0.4 \]

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 \( l \) with transition probabilities \( q^{H}_{ii'} \) for \( i, i' \in \{h, l\} \). In the data we identify health status with medical expenditures \( e_l > e_h \). 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 \( e_l = .4054, e_h = .0274 \), where health expenditures are expressed as a function of the mean wage in the economy. The implied transition matrix is given by \( q^{H}_{hh} = 0.974, q^{H}_{hl} = 0.026, q^{H}_{lh} = 0.234, \) and \( q^{H}_{ll} = 0.766 \). We assume that all the new labor market entrants are healthy so that \( q^0_h = 1 \) and \( q^0_l = 0 \).

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 25 workers. We are working on the version of the program that does not impose a constraint on the firm size. 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 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

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

\[ \text{The transition matrix in the data is computed at an annual frequency and is given by } q^{H}_{hh} = 0.93, q^{H}_{hl} = 0.07, q^{H}_{lh} = 0.63, \text{ and } q^{H}_{ll} = 0.37. \]
the literature. We chose $\phi = 0.9$ and $\sigma_\epsilon = 0.1$ so that the idiosyncratic productivity process is somewhat less persistent and more volatile than the aggregate productivity. These choices are within the (wide) range of available estimates (see, e.g., Hopenhayn and Rogerson (1993), Comin and Philippon (2005), Cooper and Haltiwanger (2006), Khan and Thomas (2008), Bachmann, Caballero, and Engel (2008)). We determine the shock values $z_i$ and the transition matrix $Q(z, \cdot)$ for a 15-state Markov chain $\{z_1, z_2, ..., z_{15}\}$ intended to approximate the postulated continuous-valued autoregression. We restrict $z_1$ and $z_{15}$ as implied by three unconditional standard deviations of $\ln(z)$ above and below the unconditional mean of the process, respectively.

Remaining Parameters. Three parameters remain to be determined: the values of non-market activity, $b$, the matching function parameter, $\chi$, and the unconditional mean of idiosyncratic productivity shocks process $\zeta$. 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.

4.2 Benchmark Results

Despite the apparent simplicity of the model it is very difficult to compute because of the discreteness which requires the use of mixed strategies at many points in the state space. We worked on the program for months and it is finally running and converging. We will have the first set of results within several weeks. Since our primary goal is to understand the properties of the US health care system these results will be interesting regardless of what they actually are. Thus, we are confident we will have something interesting to talk about in time for the conference.
Table 1: Calibrated Parameter Values.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Definition</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>the time discount rate</td>
<td>0.99</td>
</tr>
<tr>
<td>$1 - \rho$</td>
<td>the probability of a worker leaving the labor market</td>
<td>0.00625</td>
</tr>
<tr>
<td>$\eta$</td>
<td>curvature of the production function</td>
<td>0.85</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>unconditional mean of idiosyncratic productivity</td>
<td>x.xxx</td>
</tr>
<tr>
<td>$\phi$</td>
<td>persistence of idiosyncratic productivity</td>
<td>0.9</td>
</tr>
<tr>
<td>$\sigma_\epsilon$</td>
<td>st. dev. of innovations in idiosyncratic productivity</td>
<td>0.1</td>
</tr>
<tr>
<td>$\chi$</td>
<td>scale parameter of the matching function</td>
<td>x.xxx</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>elasticity of the matching function</td>
<td>0.5</td>
</tr>
<tr>
<td>$c_f$</td>
<td>firms’ flow cost of operating</td>
<td>0.0</td>
</tr>
<tr>
<td>$c$</td>
<td>cost of maintaining a vacancy</td>
<td>0.58p</td>
</tr>
<tr>
<td>$b$</td>
<td>value of non-market activity</td>
<td>x.xxx</td>
</tr>
<tr>
<td>$\tau$</td>
<td>proportional tax on labor income</td>
<td>0.4</td>
</tr>
<tr>
<td>$\kappa_{ge}$</td>
<td>administrative load schedule for employer provided insurance</td>
<td>0.08</td>
</tr>
<tr>
<td>$\kappa_{ind}$</td>
<td>administrative load for individual insurance</td>
<td>0.4</td>
</tr>
<tr>
<td>$e_t$</td>
<td>health expenditures of unhealthy workers</td>
<td>0.4054</td>
</tr>
<tr>
<td>$e_h$</td>
<td>health expenditures of healthy workers</td>
<td>0.0274</td>
</tr>
<tr>
<td>$q_{hh}$</td>
<td>the health status transition probability</td>
<td>0.93</td>
</tr>
<tr>
<td>$q_{hl}$</td>
<td>the health status transition probability</td>
<td>0.07</td>
</tr>
<tr>
<td>$q_{lh}$</td>
<td>the health status transition probability</td>
<td>0.63</td>
</tr>
<tr>
<td>$q_{ll}$</td>
<td>the health status transition probability</td>
<td>0.37</td>
</tr>
<tr>
<td>$q_{h0}$</td>
<td>fraction of new entrants who are healthy</td>
<td>1.00</td>
</tr>
<tr>
<td>$q_{l0}$</td>
<td>fraction of new entrants who are unhealthy</td>
<td>0.00</td>
</tr>
</tbody>
</table>

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

4.3 Evaluating Alternative Policies
5 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 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 one's spouse's employer. Thus introducing couples may be an important extension. An interesting question is whether large firms subsidize small firms through this channel.
References


APPENDICES

I Transition Probabilities

I.1 Firm Transition $\mu(\psi'|g)$

Consider a firm in state $(z, \psi)$ with policy $g$. After workers have been induced to leave $(g_h, g_l)$ 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 $(\psi_h, \psi_l)$ after random dismissal is given by

$$q_{\text{dis}}(\psi_h, \psi_l|g) = \begin{cases} 
\frac{\binom{g_h}{\psi_h}\binom{g_l}{\psi_l}}{\binom{g_h+g_l}{g_e}}, & \text{if } \psi_h + \psi_l = g_e, \\
0, & \text{otherwise.}
\end{cases} \quad (A1)$$

The logic behind this formula is as follows. The total number of workers before random dismissal is $g_h + g_l$, and there are $g_e$ slots. There are $\binom{g_h+g_l}{g_e}$ different ways of allocating these slots, all equally likely. The number of different ways of allocating these slots which have $\psi_h$ healthy and $\psi_l$ unhealthy workers are $\binom{g_h}{\psi_h}\binom{g_l}{\psi_l}$. Vandermonde’s identity insures that these probabilities add up to one. Let $Q_{\text{dis}}(g)$ 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 $(\psi_h, \psi_l)$ to $(\psi'_h, \psi'_l)$ in this step. The number of workers must remain the same $\psi'_h + \psi'_l = \psi_h + \psi_l$, otherwise the probability of this transition is zero. For a transition to $(\psi'_h, \psi'_l)$ it must be that the number of workers remaining in status $h$ is at least $\max\{\psi'_h - \psi_l, 0\}$, because no more than $\psi_l$ can join from status $l$. The probability of $j$ workers remaining in status $h$ is given by $B\left(j; \psi_h, q_{hh}^H\right)$. 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 $\psi'_h$ requires that exactly $\psi'_h - j$ switch from status $l$ to status $h$. The latter happens with probability $B\left(\psi'_h - j; \psi_l, q_{hl}^H\right)$. Thus

$$q_{\text{health}}(\psi_h, \psi_l; \psi'_h, \psi'_l) = \max\{\psi'_h, \psi_h\} \sum_{j=\max\{\psi'_h - \psi_l, 0\}}^{\max\{\psi'_h, \psi_h\}} B\left(j; \psi_h, q_{hh}^H\right) B\left(\psi'_h - j; \psi_l, q_{hl}^H\right).$$
Notice that this probability does not depend on the firm’s policy $g$. Let $Q^{\text{health}}$ denote the transition matrix associated with this step.

In the next step workers exit the labor market with probability $1 - \rho$. The probability of a transition from $(\psi_h, \psi_l)$ to $(\psi_h', \psi_l')$ in this step is

$$q^{\text{exit}}(\psi_h, \psi_l; \psi_h', \psi_l') = B(\psi_h'; \psi_h, \rho)B(\psi_l'; \psi_l, \rho).$$

Let $Q^{\text{exit}}$ 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 $(\psi_h, \psi_l)$ to $(\psi_h', \psi_l')$ is

$$q^{\text{vac}}(\psi_h, \psi_l; \psi_h', \psi_l') \equiv B(\psi_h' + \psi_l' - \psi_h - \psi_l; g_v, q(\theta))B\left(\frac{\psi_h' - \psi_h; \psi_h' + \psi_l' - \psi_h - \psi_l, m_h}{m_h + m_l}\right).$$

The first term captures that out of $g_v$ vacancies it must be that $\psi_h' + \psi_l' - \psi_h - \psi_l$ 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 $\frac{m_h}{m_h + m_l}$. Let $Q^{\text{vac}}(g, q(\theta))$ denote the transition matrix.

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

$$\mu(\psi'|g) = \left[Q^{\text{vac}}(g, q(\theta)) \cdot Q^{\text{exit}} \cdot Q^{\text{health}} \cdot Q^{\text{dis}}(g)\right]_{\psi'}$$

where $[Q]_{\psi'}$ extracts the element of vector $Q$ corresponding to the workforce $\psi'$.

### I.2 Worker Transition $\mu_{ii'}(\psi'|g)$

We derive $\mu_{hh}(\psi'|g)$, 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 $(g_h, g_l)$ 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 \((\hat{\psi}_h, \hat{\psi}_l)\) after random dismissal is given by

\[
q^{\text{dis}}_{hh}(\hat{\psi}_h, \hat{\psi}_l | g) = \begin{cases} 
\frac{g_{h-1}}{g_e}, & \text{if } \hat{\psi}_h + \hat{\psi}_l = g_e, \\
0, & \text{otherwise.}
\end{cases}
\] (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 \(g_h + g_l - 1\), and there are \(g_e - 1\) remaining slots. There are \(\binom{g_h + g_l - 1}{g_e - 1}\) different ways of allocating these slots, all equally likely. To end up at \((\hat{\psi}_h, \hat{\psi}_l)\) 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 \(\hat{\psi}_l\) unhealthy workers are \(\binom{g_h - 1}{\hat{\psi}_h - 1}\). Let \(Q^{\text{dis}}_{hh}(g)\) denote the vector of these probabilities.

Next workers draw their new health status. We compute the probability of a transition from \((\psi_h, \psi_l)\) to \((\psi'_h, \psi'_l)\) The number of workers must remain the same \(\psi_h + \psi_l = \psi'_h + \psi'_l\), 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 \((\psi'_h, \psi'_l)\) it must be that the number of workers remaining in status \(h\) is at least \(\max\{\psi'_h - \psi_l, 1\}\), because no more than \(\psi_l\) can join from status \(l\), and we already condition on one healthy worker staying healthy. The probability of \(j\) workers remaining in status \(h\) is given by \(B(j - 1; \psi_h - 1, q_{hh}^H)\). If \(j\) workers remain in status \(h\), the transition to \(\psi'_h\) requires that exactly \(\psi'_h - j\) switch from status \(l\) to status \(h\). The latter happens with probability \(B(\psi'_h - j; \psi_l, q_{hl}^H)\). Thus

\[
q^{\text{health}}_{hh}(\psi_h, \psi_l; \psi'_h, \psi'_l) = \sum_{j=\max(\hat{\psi}_h - \hat{\psi}_l, 1)}^{\max\{\hat{\psi}_h, \hat{\psi}_l\}} B(j - 1; \hat{\psi}_h - 1, q_{hh}^H) B(\hat{\psi}_h - j; \hat{\psi}_l, q_{hl}^H).
\]

Let \(Q^{\text{health}}_{hh}\) denote the associated transition matrix.

Next workers exit the labor force with probability \((1 - \rho)\). We condition on the worker under consideration remaining in the labor market. The probability of a transition from \((\psi_h, \psi_l)\) to \((\psi'_h, \psi'_l)\) in this step is

\[
q^{\text{exit}}_{hh}(\psi_h, \psi_l; \psi'_h, \psi'_l) = B(\psi'_h - 1; \psi_h - 1, \rho) B(\psi'_l; \psi_l, \rho).
\]
Let $Q^\text{exit}_{hh}$ 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 steps, we have Combining these transitions, the distribution $\mu(\psi'|g)$ is given by

$$
\mu(\psi'|g) = [Q^\text{vac}(g, q(\theta)) \cdot Q^\text{exit}_{hh} \cdot Q^\text{health}_{hh} \cdot Q^\text{dis}_{hh}(g)]_{\psi'}$

I.3 Searching Worker Transition $\mu^z_i(\psi', \psi')$

A searching worker who makes contact is randomly allocated to a vacancy. Let $(z, \psi)$ 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 $\gamma(g|z, \psi)$, in which case it has $g_v$ vacancies. Thus the searcher is matched with a firm implementing policy $g$ with probability

$$
q^\text{match}(g) = \frac{\sum_{(z,\psi)\in S} g_v \gamma(g|z, \psi) \mu(z, \psi)}{\sum_{g'\in G} \sum_{(z,\psi)\in S} g'_v \gamma(g'|z, \psi) \mu(z, \psi)}.
$$

If matched with a firm implementing policy $g$, the distribution of that firm’s workforce after random dismissal is $Q^\text{dis}(g)$, and after health status changes and exit from the labor market it is $Q^\text{exit} \cdot Q^\text{health} Q^\text{dis}(g)$. 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 $i' = l$ is analogous. If a firm has workforce $(\psi_h, \psi_l)$ before vacancies are filled and follows policy $g$, then the probability to arrive at $(\psi'_h, \psi'_l)$ is

$$
q^\text{vac}_h(\psi_h, \psi_l; \psi'_h, \psi'_l|g) \equiv B\left(\psi'_h + \psi'_l - \bar{\psi}_h - \bar{\psi}_l - 1; g_v - 1, q(\theta)\right)
\cdot B\left(\psi'_h - \bar{\psi}_h - 1; \psi'_h + \psi'_l - \bar{\psi}_h - \bar{\psi}_l - 1, \frac{m_h}{m_h + m_l}\right).
$$

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

$$
\mu_{h}^{s}(z', \psi') = \sum_{g \in G} [Q_{h}^{vac}(g) \cdot Q^{exit} \cdot Q^{health} \cdot Q^{dis}(g)]_{\psi'} q_{zz'} q_{match}(g)
$$

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 $J(\cdot), V_{h}(\cdot), V_{l}(\cdot), V_{h}^s, V_{l}^s, \mu(\cdot), m_{h}, m_{l},$ and $\theta$ consistent with that policy. Second, for any $\{J(\cdot), V_{h}(\cdot), V_{l}(\cdot), V_{h}^s, V_{l}^s, \mu(\cdot), m_{h}, m_{l}, \theta\}$ we can use equation (7) to compute a set of policies which are optimal. Combining these two mappings, we get a correspondence $\Omega$ 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 $\gamma^{k}(\cdot|\cdot; \bar{\gamma}(\cdot|z, \psi))$ denote the policy given by

$$
\gamma^{k}(\cdot|z', \psi'; \bar{\gamma}(\cdot|z, \psi)) = \begin{cases} 
\gamma^{k}(\cdot|z', \psi') & \text{for all } (z', \psi') \neq (z, \psi), \\
\bar{\gamma}(\cdot|z, \psi) & \text{for } (z', \psi') = (z, \psi).
\end{cases}
$$

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

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

$$
\Omega_{(z, \psi)}[\gamma(\cdot|\cdot)] = \{\bar{\gamma}(\cdot|z, \psi) \in \Gamma|\bar{\gamma}(\cdot, \cdot) \in \Omega[\gamma(\cdot|\cdot)]\}.
$$

This is the sets of mixed policies that are optimal for the point in the state space $(z, \psi)$ 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 $(z^{k}, \psi^{k}) \in S$ at random.
2. Given the policy function $\gamma_k(\cdot|\cdot)$, find a mixed policy $\tilde{\gamma}(\cdot|z^k,\psi^k)$ such that

$$
\gamma_k(\cdot|\cdot; \tilde{\gamma}(\cdot|z^k,\psi^k)) \in \Omega(\gamma_k(\cdot|\cdot; \tilde{\gamma}(\cdot|z^k,\psi^k))).
$$

This step is implemented using a heuristic algorithm.

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

III Accounting

In the firm transition, if the policy is $(g_h, g_l, g_e, g_v)$, where can we end up? Employment cannot exceed $g_e + g_v$, since $g_e$ is the number of workers at the time of production, and at most $g_v$ workers can join, which happens if all vacancies are filled. Employment could be zero if all workers exit the labor market and no vacancies are filled. It is possible for all workers to be healthy, which occurs if all unhealthy workers become healthy, all healthy workers stay healthy, and vacancies are filled only with healthy workers. Similarly, it could be that all workers are unhealthy. So $\psi_h \in \{0, g_e + g_v\}$. If there are $\psi_h$ healthy workers, then the number of unhealthy workers must be $\psi_l \in \{0, g_e + g_v - \psi_l\}$ So the total number of non-zero entries is

$$
\sum_{k=0}^{g_e + g_v} (g_e + g_v - k)
$$

IV Modeling Assumptions

Dismissal. There is a mediator, and workers report their health status to the mediator and receive recommendations about leaving or staying. Suppose there are 6 unhealthy workers and 10 healthy workers. Suppose that if all the healthy workers stay and 5 unhealthy stay, then the remaining unhealthy worker is willing to leave, while all health workers are willing to stay. If all workers truthfully report their health status, then he randomly chooses one among the unhealthy and recommends leaving. The question is what the mediator should do if an unhealthy worker deviates and reports being healthy.

In the text we allow firms to dismiss any desired number of a given health status, as long as leavers would not want to deviate and stay. We assume that within a given
health status leavers are chosen at random. How can the firm implement this if it does not know the worker’s health status, and also worker’s do not know each other’s health status?

For concreteness, suppose the firm has 20 healthy workers and 5 unhealthy workers, and it wants 2 unhealthy workers to leave. Consider the decision of an unhealthy worker. Suppose that, given that 20 healthy workers and 3 other unhealthy workers stay, then this unhealthy does not benefit strictly from staying. But if 20 healthy and 2 other unhealthy stay, then he does not benefit strictly from leaving. Similarly, suppose that, given that 19 other healthy workers and 3 unhealthy workers stay, a healthy worker does not benefit strictly from leaving. This is precisely the situation in which we assume that (20, 3) is a feasible choice for the firm.

The firm can implement this by setting up an urn with five envelopes, three of which are empty and two of which contain pink slips. The workers, unobserved by the firm and other workers, choose whether to take an envelope from the urn, and whether to stay or leave. The following is an equilibrium. Healthy workers do not take an envelope. Unhealthy workers do. Unhealthy workers with a pink slip leave, all others stay.