Essays on Economies with Heterogeneous Labor

by

Brandon Charles Lehr

B.A., University of California at Berkeley (2005)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Economics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2010

Author ....
Department of Economics
June 22, 2010

Certified by ....
Peter Diamond
Institute Professor and Professor of Economics
Thesis Supervisor

Certified by ....
Ivan Werning
Professor of Economics
Thesis Supervisor

Accepted by ....
Esther Duflo
Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics
Chairman, Departmental Committee on Graduate Studies

Essays on Economies with Heterogeneous Labor

by

Brandon Charles Lehr

Submitted to the Department of Economics
on June 22, 2010, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Economics

Abstract

In this thesis, I study two different economies that consist of heterogeneous labor. By allowing for differences among individuals where previous analyses restricted attention to homogeneous labor, I am able to understand the impact of such a consideration on issues of optimal policy and potential equilibria.

The first chapter, *Optimal Social Insurance with Individual Private Insurance and Moral Hazard*, characterizes optimal social insurance in an economy where competitive firms also provide insurance to workers facing uncertain outcomes. An ex-ante heterogeneous population of workers exerts effort to increase the likelihood of high outcome events. This chapter is novel in its joint consideration of two sources of heterogeneity, two potential sources of insurance, and an endogenous ex-post distribution of outcomes. The introduction of ex-ante heterogeneity in the presence of optimal private insurance changes the optimal prescription for social insurance away from zero. Moreover, the relative source of the variation in outcomes due to ex-ante heterogeneity and ex-post shocks plays a significant role in the welfare loss associated with setting optimal social insurance without recognizing the presence of private insurance.

The second chapter, *Efficiency Wages with Heterogeneous Agents*, builds a model of efficiency wages with heterogeneous workers in the economy who differ with respect to their disutility of labor effort. In such an economy, two types of pure strategy symmetric Nash equilibria in firm wage offers can exist: a no-shirking equilibrium in which all workers exert effort while employed and a shirking equilibrium in which within each firm some workers exert effort while others shirk. The type of equilibrium that prevails in the economy depends crucially on the extent of heterogeneity among the workers. In addition, it is shown that the characterization of the economy is independent of allowing for variable labor hours and the subsequent adverse selection problem it introduces, as there does not exist a pure strategy symmetric separating Nash equilibrium.

Finally, in the third chapter I correct the proof of the main proposition in the analysis of an efficiency wage model with a continuum of heterogeneous agents constructed by Albrecht and Vroman (1998).

Thesis Supervisor: Peter Diamond
Title: Institute Professor and Professor of Economics

Thesis Supervisor: Ivan Werning
Title: Professor of Economics

Acknowledgments

I am grateful for the advice and intellectual guidance provided by my advisors Peter Diamond and Ivan Werning. I have also received helpful suggestions from numerous faculty and students at MIT, especially Emmanuel Saez, Jim Poterba, and Amy Finkelstein. Additionally, I want to acknowledge Raj Chetty for introducing me to the field of public economics and encouraging me to pursue my early research agenda. Finally, I am indebted to the personal guidance and support provided by Brandon Andrew and my parents. Thank you.

Contents

List of Figures

Chapter 1

Optimal Social Insurance with Individual Private Insurance and Moral Hazard

1.1 Introduction

Social insurance provides the government with the ability to insure individuals against risk and also to redistribute across agents in the economy. If we accept that in a market economy the social planner cannot directly control or exclude the provision of private insurance, it is important for the government to internalize the endogenous response of private insurance markets and individual behavior to social policy. Failure to recognize the presence of privately provided insurance can lead to incorrect prescriptions for optimal policy that may result in substantial welfare losses. This chapter therefore takes up the task of characterizing analytically and numerically optimal social insurance in the presence of endogenously supplied private insurance.

This chapter builds a model of the economy in which a heterogeneous population of workers differ with respect to productivity and also face an ex-post shock to earnings. Workers exert effort to increase the likelihood of a high shock. This effort is unobservable to insurers, both private and public. The important assumption, however, is that private insurers have better information than the government and in particular can observe the productivity of each agent. Thus, private insurance insures risk for a single agent, while social insurance can redistribute across agents.

In the first part of the chapter we obtain analytical results to characterize the impact of private insurance markets on the provision of optimal social insurance. In order to do so, we assume that the realization of the ex-post shock is observable to the social planner. A natural interpretation is one in which the negative shock is that of entering unemployment. In that case, private insurance can take the form of a severance package from firms or private unemployment insurance as provided by a union, for example. We could also apply this setting to natural disaster, disability, or health insurance. In each of these cases, agents exert effort to prevent a low outcome event and actuarially fair private insurance helps to smooth consumption in case of an accident. The private insurers offer type-contingent insurance contracts, whereas the government only observes the realization of a low outcome event and provides the same social insurance contract to everyone.

In the second part of the chapter we allow for more general shocks that do not need to be observable by the government. In particular, ex-post shocks take the form of productivity shocks and private insurance comes in the form of a compressed wage structure. This insurance is equivalent to the solution of a firm solving a standard moral hazard problem. The additional social insurance is enacted via a linear income tax schedule. This extends the standard optimal income taxation literature stemming from the "hidden skill" model of Mirrlees (1971)[24] in which wages are exogenously determined by the productivity of each agent. Instead, firms offer contracts to workers which impact realized earnings and are conditional on the tax program in the economy.

This chapter characterizes the optimal social insurance in the presence of individual private insurance and numerically simulates the importance of such considerations. The main analytical result is that the desirability of social insurance in this model depends on the covariance between income and risk. In particular, if higher earners have a lower probability of bad shocks relative to the poor, the case for social insurance is strengthened. This result is analogous to the conclusions in Rochet (1991)[27] and Cremer and Pestieau (1996)[8], despite the more complex setting in this chapter of allowing for explicit moral hazard concerns.

In many cases the optimal level of social insurance is reduced by taking into account the presence of private insurance. Chetty and Saez (2009)[7], Golosov and Tsyvinski (2007)[15], and Kaplow (1991)[20] make this point theoretically. Empirical work by Cutler and Gruber (1996)[10][11] and Gruber and Simon (2008)[17] suggests that crowdout is significant, with the implication that less social insurance can be welfare improving.[1] There exist parameterizations of the model in this chapter, however, in which social insurance should be higher with the presence of private insurance. This is due to the impact of insurance on effort, which in turn determines the ex-post distribution of earnings. Private insurance reduces effort, increasing the likelihood of low states of nature, and increasing the need for social insurance in some settings.

Moreover, the difference between optimal social insurance with and without private insurance depends on the ex-ante variance of productivity types. When ex-ante heterogeneity is large, firms provide little in the way of insurance across the economy, so the optimal social insurance is largely unaffected by the presence or absence of private insurance. However, when ex-ante heterogeneity is small, private insurance for a single individual, who is largely representative of the average worker in the economy, obviates the need social insurance across individuals.

The analysis here extends a previous literature on optimal insurance. There exists a class of models in which there is only a single source of heterogeneity. For example, agents are ex-ante identical, but experience exogenous "luck," which introduces uncertainty and taxation therefore serves as social insurance (e.g. Diamond, Helms, and Mirrlees (1980)[12], Eaton and Rosen (1980)[13], and Varian (1980)[30]). Kaplow (1991)[20] allows for both private and social insurance, where unobservable actions impact the probability of a high outcome event, but all agents are ex-ante identical. Chetty and Saez (2009)[7] build a similar model with a single source of earnings heterogeneity, but they provide expressions for the welfare gains from government inter-

[1] An exception, Finkelstein (2004)[14] shows in the health insurance context that there is little impact of the partial insurance program of Medicare on the private insurance market.

vention when private insurance is not optimal. Another literature considers optimal social insurance in the presence of two sources of heterogeneity due to differences in earnings ability and exogenous ex-post shocks to income. Mirrlees (1990)[25] is able to characterize the optimal linear tax in such a setting, while Rochet (1991)[27] and Cremer and Pestieau (1996)[8] consider jointly optimal income taxation and social insurance without moral hazard.

The paper most closely related to the analysis of this chapter is Boadway et al. (2003)[3] in which there also exist two sources of heterogeneity and both social and private insurance in the presence of moral hazard. Their application is the health insurance market and they allow for the government to offer both a linear tax system and an insurance program, as opposed to the analysis here in which taxation is the only government tool, which acts to potentially both insure and redistribute. Moreover, private insurance does not affect taxable earnings in their analysis. This may be justifiable in the health care context, but is certainly not a reasonable assumption when considering more general insurance in the form of wage compression for agents facing uncertain productivity outcomes. The analysis in this chapter also has the advantage of characterizing the welfare impact of increasing (decreasing) social insurance with expressions that are in terms of observable elasticities (in the spirit of Chetty (2009)[6]). In addition, this chapter addresses the question of quantifying welfare losses when a social planner fails to recognize the scope of the private insurance market.

The chapter is structured as follows. Section 2 presents the setup of the model. Section 3 presents a simple case in which agents are ex-ante identical as a benchmark for the general analysis in section 4. Section 5 provides numerical simulations of the optimal social insurance and Section 6 concludes. Supplementary proofs and figures are relegated to Appendix A.

1.2 The Model

The economy consists of workers, indexed by productivity θ, with a distribution given by $f(\theta)$. Each worker inelastically supplies labor hours, but a variable level of effort $a \in [0,1]$. A worker's effort results in either a high or low state relative to the individual's innate productivity. In particular, the possible earnings for an individual of type θ is assumed to either be $M(\theta)$ or $m(\theta)$ where $M(\theta) > m(\theta)$ for all θ. The probability of the high outcome is given by a. Workers are risk averse and have a standard separable utility function over consumption, c, and effort given by

$$U(c, a) = u(c) - h(a) \tag{1.1}$$

where $u(\cdot)$ is continuously differentiable, increasing, and concave and $h(\cdot)$ is differentiable, increasing, and convex, with $h'(0) = 0$ and $h'(1) = \infty$.

Workers choose effort to maximize expected utility. Denote consumption in the two possible states for a worker of type θ by $c_1(\theta)$ and $c_0(\theta)$ for the high and low realizations, respectively. We assume throughout that $c_1(\theta) > c_0(\theta)$ so that there is never full insurance. The first order condition for the agent's choice of effort is given by

$$u(c_1) - u(c_0) = h'(a) \tag{1.2}$$

One interpretation of this economy is that in which the low state of nature is unemployment and the high state is employment. Since the states of nature are observable to the government, while the worker's innate productivity is not, the government provides insurance to all workers in the form of a tax, t, when employed and a benefit, b, when unemployed.[2] We also assume the existence of actuarially fair private insurance contracts that can supplement the public insurance program.[3] The

[2]Note that the government does not use earnings data to infer an individual's type, either because the earnings functions ($M(\theta)$ and $m(\theta)$) are not invertible, or the administrative cost of truthfully collecting such data outweighs the gain of constructing a more sophisticated social insurance program.

[3]The implicit assumption is that firms are large so that the law of large numbers implies firm

agent's innate productivity is public information to the firm, while the effort supplied by the agent is unobservable and noncontractable. Thus, private insurance contracts are type-dependent and consist of a tax, $t_p(\theta)$, and benefit, $b_p(\theta)$ for a worker of type θ.[4] The benefit in the low state provided by the private insurance contract can be understood as a severance payment or privately provided unemployment insurance benefit.

The information structure of this problem is one in which firms have better information than the government. Although the assumption of observable type may not be wholly realistic, it allows us to focus on the moral hazard problem of the firm over the adverse selection issue. The market failure induced by adverse selection introduces a role for the government and this analysis is concerned with understanding whether or not social insurance is beneficial even without the informational problem of adverse selection in the private market.[5] Moreover, it is reasonable that firms would have better information about the types of their own workers relative to the tax authority.

In the following sections, we characterize optimal social insurance in the presence of optimal private insurance contracts under two cases.[6] First, we assume a degenerate ex-ante distribution of types in the economy so that all workers are equally productive. The only heterogeneity in the population is due to the ex-post productivity realization. This case provides a useful benchmark in recognizing that with only one dimension of heterogeneity, the government cannot improve upon the private insurance contracts. Next, we turn to the general case in which there are two sources of heterogeneity and characterize the impact of such ex-ante heterogeneity on optimal insurance as determined by a utilitarian social planner.

---

profits are in fact equal to zero, and not simply zero in expectation. This allows for welfare to depend simply on the utility of the consumers in the economy.

[4]As in standard problems of moral hazard, the private insurer perfectly predicts the effort level of each agent, but the incentive scheme is necessary to induce the optimal effort supply.

[5]Boadway et. at. (2006)[4] note that the case for social insurance can be strengthened in the presence of both adverse selection and moral hazard.

[6]By "optimal", we assume throughout the analysis a utilitarian social planner. The allocations in the private economy are Pareto efficient and the role of the social planner is to redistribute to achieve a more equitable allocation on the Pareto frontier.

1.3 Ex-Ante Homogeneous Agents

1.3.1 Optimal Social Insurance without Private Insurance

Consider a population of identical agents with productivity θ. Let the government be the only provider of insurance in this economy, with a tax, q, and a benefit, s. Thus, consumption of agents is given by $c_1 = M(\theta) - q$ and $c_0 = m(\theta) + s$. The government must break even, imposing the constraint that $aq = (1-a)s$. The government thus chooses s to maximize

$$W = au\left(M(\theta) - \frac{1-a}{a}s\right) + (1-a)u(m(\theta)+s) - h(a) \tag{1.3}$$

Since a maximizes utility, by the envelope theorem we have that the first order condition with respect to s is:

$$\begin{aligned} 0 = \frac{dW}{ds} &= (1-a)u'(c_0) - au'(c_1)\left(\frac{1-a}{a} + s\frac{d\left(\frac{1-a}{a}\right)}{ds}\right) \\ &= (1-a)u'(c_1)\left(\frac{u'(c_0)-u'(c_1)}{u'(c_1)} - \frac{\varepsilon_{1-a,s}}{a}\right) \end{aligned} \tag{1.4}$$

where we define $\varepsilon_{1-a,s}$ to be the elasticity of the probability of the low consumption event with respect to the benefit, s. For a nonzero level of effort, an increase in the benefit reduces effort, as seen by the worker's first order condition in (1.2). In order to satisfy the rule for optimal social insurance provision in (1.4) we must have that $\varepsilon_{1-a,s} > 0$, so it follows that $s > 0$. Equation (1.4) yields the following proposition.

Proposition 1 The optimal social insurance contract with ex-ante homogeneous agents and no private insurance is characterized by

$$\frac{u'(c_0)-u'(c_1)}{u'(c_1)} = \frac{\varepsilon_{1-a,s}}{a} \tag{1.5}$$

Proof Immediate from (1.4). ■

The left hand side of (1.5) measures the marginal value of insurance as the difference between marginal utilities across the two states. The right hand side measures the marginal cost of insurance via the behavioral distortion. At the optimum, these must be equal. This analysis provides a benchmark for understanding how the introduction of private insurance impacts the role of social insurance.

1.3.2 Optimal Social Insurance in the Presence of Optimal Private Insurance

With the introduction of private insurance, denote the government's social insurance contract by (b, t). Private insurers and workers take this as given. Firms respond by setting the optimal private insurance contract, (b_p, t_p). Expected firm profits must be zero, thus imposing the constraint that $at_p = (1-a)b_p$. It is also useful to define the crowdout parameter, $r = -\frac{db_p}{db}$, to measure the response of the private insurers to the government insurance contract. When $r = 0$ there is no crowdout and with $r = 1$, there is 100% crowdout of private insurance. In this setting with ex-ante homogeneous agents, we have the following result, which is analogous to the analysis in Chetty Saez (2009)[7].

Proposition 2 When private insurance is set optimally for a population of ex-ante homogeneous agents,

1. The optimal social insurance contract is $b = 0$

2. The marginal effect on welfare of an increase in the benefit b is given by

$$\begin{aligned}\frac{dW}{db} &= -\frac{1-a}{a}u'(c_1)(1-r)\frac{b}{b+b_p}\varepsilon_{1-a,b+b_p} \\ &\equiv -\frac{1-a}{a}u'(c_1)\varepsilon_{1-a,b} \end{aligned} \tag{1.6}$$

   where $\varepsilon_{1-a,b}$ is the elasticity of the low probability event with respect to b, taking into account the response of b_p to b.

Proof

Optimal Private Contract. Firms, taking (b,t) as given choose b_p to maximize

$$W = au\left(M(\theta) - t - \frac{1-a}{a}b_p\right) + (1-a)u(m(\theta)+b+b_p) - h(a) \tag{1.7}$$

Using the envelope theorem for the agents' optimal choice of a, we have that

$$\begin{aligned} 0 = \frac{dW}{db_p}|_{b,t} &= (1-a)u'(c_0) - au'(c_1)\left(\frac{1-a}{a} + b_p\frac{d\left(\frac{1-a}{a}\right)}{db_p}\right) \\ &= (1-a)u'(c_1)\left(\frac{u'(c_0)-u'(c_1)}{u'(c_1)} - \frac{\varepsilon_{1-a,b_p}|_{b,t}}{a}\right) \end{aligned} \tag{1.8}$$

We can define $s = b+b_p$ to be the total insurance provided to individuals and note that $\varepsilon_{1-a,b_p}\big|_{b,t} = \frac{d(1-a)}{db_p+db}\frac{b_p}{1-a} = \frac{d(1-a)}{db_p+db}\frac{b_p}{1-a}\frac{b+b_p}{b+b_p} = \varepsilon_{1-a,s}\frac{b_p}{b+b_p}$. The optimal private insurance contract, conditional on (b,t), is therefore characterized by

$$\frac{u'(c_0)-u'(c_1)}{u'(c_1)} = \frac{\varepsilon_{1-a,s}}{a}\frac{b_p}{b+b_p} \tag{1.9}$$

Note that in the case in which the government is not providing any insurance ($b=0$), the formula for optimal private insurance reduces to the same rule in (1.5) for optimal social insurance without private insurance.

Social Insurance in the Presence of Optimized Private Insurance. The government chooses benefit b and tax $t = \frac{1-a}{a}b$, to maximize social welfare, taking into account b_p set as in (1.9). To simplify the problem, note that $c_1 = M(\theta) - \frac{1-a}{a}s$ and $c_0 = m(\theta)+s$ where again $s = b + b_p(b)$. With a change of variables, it is therefore equivalent for the government to choose s instead of b directly. In particular, from (1.7)

$$W = au\left(M(\theta) - \frac{1-a}{a}s\right) + (1-a)u(m(\theta)+s) - h(a)$$

and

$$\frac{dW}{db} = \frac{dW}{ds}\frac{ds}{db} = (1-r)\frac{dW}{ds}$$
$$= (1-r)(1-a)u'(c_1)\left(\frac{u'(c_0)-u'(c_1)}{u'(c_1)} - \frac{\varepsilon_{1-a,s}}{a}\right)$$

where the expression for $\frac{dW}{ds}$ follows from (1.4). Plugging in the expression from (1.9), we have that

$$\frac{dW}{db} = (1-r)(1-a)u'(c_1)\left(\frac{\varepsilon_{1-a,s}}{a}\frac{b_p}{b+b_p} - \frac{\varepsilon_{1-a,s}}{a}\right)$$
$$= -(1-r)\frac{1-a}{a}u'(c_1)\varepsilon_{1-a,s}\frac{b}{b+b_p}$$

This expression establishes that $\frac{dW}{db}(b=0)=0$. To show that $b=0$ is the global maximum, it is sufficient to establish that $r<1$. If an increase in b were exactly offset by a decrease in b_p, maintaining a constant level of total insurance, effort would be unchanged and we would violate (1.9). Thus, b_p must fall by less than the increase in b to satisfy (1.9). Moreover, we can simplify the above expression by noting that $\varepsilon_{1-a,s} = \frac{d(1-a)}{ds}\frac{s}{1-a}\frac{b}{b}\frac{db}{db} = \varepsilon_{1-a,b}\frac{b+b_p}{b}\frac{1}{1-r}$. Therefore, we have

$$\frac{dW}{db} = -\frac{1-a}{a}u'(c_1)\varepsilon_{1-a,b}$$

as claimed in (1.6). ■

Proposition 2 establishes that when individuals are ex-ante homogeneous, there is no role for social insurance in the presence of optimally provided private insurance. To better understand the intuition for this result, first note the optimal private insurance as characterized in expression (1.9). We see that in the case in which there is no government intervention ($b=0$), (1.9) reduces to (1.5) and firms provide the same level of optimal social insurance in the absence of private insurance contracts. The effect on welfare of increasing b at zero equals the deadweight burden of greater taxation, implying that the government should do nothing. Hence, the optimal level of insurance can be equivalently provided exclusively by the private sector or the social planner.

The impact of social policy on welfare, $\frac{dW}{db}$, is expressed in terms of two different elasticities in (1.6). The first is an expression in terms of the behavioral elasticity with respect to total insurance, $s = b + b_p$. This elasticity, $\varepsilon_{1-a,s}$, however, is less empirically useful since we do not observe the total change in insurance and thus need an estimate of the crowdout parameter, r. The second expression employs the fact that $\varepsilon_{1-a,b} = (1-r)\frac{b}{b+b_p}\varepsilon_{1-a,s}$. This expression relates the observable elasticity of behavior with respect to policy to the more fundamental elasticity of behavior with respect to total insurance, which may not be observable.[7] The intuition for this identity is that an increase in b leads to a smaller change in s via two channels. The first is that an increase in b leads to crowdout of b_p, so we must scale the fundamental elasticity by $(1-r)$. In addition, there is a mechanical effect of adjusting the impact of b on s by the the relative magnitude of b to s. Hence, when b is small relative to the private insurance, b_p, the proportional effect of b on s is diminished. This second channel mechanically implies by the definition of an elasticity that $\varepsilon_{1-a,b}(b=0) = 0$, verifying the intuition for $\frac{dW}{db}(b=0) = 0$ as explained above.

When the government provides positive social insurance in the presence of optimal private insurance, there is a negative marginal effect on welfare. The intuition for this result is as follows. Given some $b > 0$, suppose firms topped up the insurance by setting b_p such that $b + b_p$ were equal to the optimal level of social insurance in the absence of private contracts. Such a private insurance contract will satisfy (1.5) by definition, but will then not satisfy (1.9) as required for optimality. In fact, private insurers perceive a lower marginal cost of providing insurance given b, and will increase b_p so as to satisfy (1.9). Thus, total insurance is greater than would be optimal with a single social insurance contract. This additional insurance reduces effort, which in turn requires greater taxes from the government to balance its budget. The failure of private insurers to internalize their impact on the government's budget

[7] I adopt the terminology of the elasticity with respect to s as the "fundamental elasticity" from the analysis in Chetty and Saez (2009)[7].

constraint leads to a strict welfare loss.

An analogous argument explains why $\frac{dW}{db}(b < 0) > 0$. Suppose the government were to exacerbate the difference in earnings levels in the two states of nature by setting $b < 0$. It would not be optimal for private insurers to compensate for the lack of social insurance by setting a total level of insurance equal to the optimal insurance when provided by a single source. In order to satisfy (1.9), private insurers would have to decrease b_p. Thus, when $b < 0$, total insurance is less than when there is a single insurer. This decreased insurance increases effort, allowing the government to lower taxes for the same level of b. Since private insurers do not internalize this impact of insurance and its induced change in effort on the government budget constraint, there is a welfare improvement by increasing b. Note that an increase in b crowds out private insurance as before so that the total level of insurance increases, but by less than the change in b. This increase in s reduces effort. However, somewhat counterintuitively, $\varepsilon_{1-a,b}(b < 0) < 0$. This is simply because when working with elasticities as opposed to derivatives, an increase in b when b is negative corresponds to a percentage decrease in b.

Chetty and Saez (2009)[7] enrich a similar environment by analyzing the optimal social insurance contract in the presence of not necessarily optimized private insurance. Instead, this chapter is focused on understanding how optimal social insurance is affected by the introduction of two sources of heterogeneity. In fact, with ex-ante heterogeneity, we can break the result of Proposition 2. We turn to this analysis in the following section.

1.4 Ex-Ante Heterogeneous Agents

1.4.1 Optimal Social Insurance without Private Insurance

We now allow for a general distribution of types in the population. With the government as the only provider of insurance, the insurance acts to insure individual workers and redistribute across types. The government must again break even, so a contract consisting of a benefit s and tax q must satisfy $q\overline{a} = s(1-\overline{a})$ where $\overline{a} \equiv \int a(\theta)f(\theta)d\theta$. We also define $\varepsilon_{1-\overline{a},s}$ as the elasticity of the average probability of low events occurring across agents with respect to the benefit level, s. We are now ready to state the analogous result to Proposition 1 as follows.

Proposition 3 The marginal effect of an increase of a benefit s on welfare is given by:

$$\frac{dW}{ds} = E((1-a)u'(c_0)) - E(au'(c_1))\frac{1-\overline{a}}{\overline{a}}\left(1 + \frac{\varepsilon_{1-\overline{a},s}}{\overline{a}}\right) \tag{1.10}$$

and the optimal social insurance contract with no private insurance is characterized by

$$\frac{\overline{a}}{1-\overline{a}} \frac{E((1-a)u'(c_0))}{E(au'(c_1))} - 1 = \frac{\varepsilon_{1-\overline{a},s}}{\overline{a}} \tag{1.11}$$

Proof The government chooses s to maximize

$$W = \int \left(a(\theta)u\left(M(\theta) - \frac{1-\overline{a}}{\overline{a}}s\right) + (1-a(\theta))u(m(\theta)+s) - h(a(\theta))\right) f(\theta)d\theta$$

Since a maximizes utility, by the envelope theorem we have that the first order condition with respect to s is:

$$\begin{aligned}\frac{dW}{ds} &= E((1-a)u'(c_0)) - E(au'(c_1))\left(\frac{1-\overline{a}}{\overline{a}} + s\frac{d\left(\frac{1-\overline{a}}{\overline{a}}\right)}{ds}\right) \\ &= E((1-a)u'(c_0)) - E(au'(c_1))\frac{1-\overline{a}}{\overline{a}}\left(1 + \frac{\varepsilon_{1-\overline{a},s}}{\overline{a}}\right)\end{aligned}$$

Setting $\frac{dW}{ds} = 0$, we can easily solve for (1.11). ■

The addition of ex-ante heterogeneity does not change the fundamental notion that optimal social insurance should equate the expected marginal benefit, as measured by the relative difference between the marginal utilities across states, and the marginal cost, as measured by the average behavioral response. Without heterogeneity, $\overline{a} = a$, the expectation operators drop out, and (1.11) reduces to (1.5). A notable difference is that the expectations of the marginal utilities are weighted by the relative probability of the high or low state occurring for a given type of individual. This is because each type of worker chooses a different effort level and thus realizes different states of nature with a different probability distribution. When those with the highest consumption experience the least risk (high effort), the left hand side of (1.11), which measures the marginal benefit of social insurance, is greater. Thus, the case for social insurance is greater when risk and income are negatively correlated. This confirms the same intuition from other models in which there exists heterogeneity with respect to ability and risk (Rochet (1991)[27] and Cremer and Pestieau (1996)[8]).

Such endogeneity of the distribution is in contrast to the analysis of optimal taxation as social insurance in Mirrlees (1990)[25]. Mirrlees allows for workers to be ex-ante different, but then experience a productivity shock that is independent of ability and labor choice. With a fixed distribution of productivity realizations, he is able to provide a simple approximation to the optimal tax rate as a linear combination of the ex-ante and ex-post variances. The simplicity of the characterization in terms of only the variances of the different sources of heterogeneity relies, however, on the assumption that both variances and tax rates are small. This chapter does not impose such restrictions.

We now turn to the fully general case in which two sources of insurance work to smooth consumption within and across individuals.

1.4.2 Optimal Social Insurance in the Presence of Optimal Private Insurance

The agents in this economy can be individually insured by the firm at which they are employed. The government can augment this insurance with its own contract, which has the added benefit of being able to redistribute across agents. Thus, firms offer contracts $(b_p(\theta), t_p(\theta))$ to each worker in which $a(\theta)t_p(\theta) = (1 - a(\theta))b_p(\theta)$. Firms and agents take the government contract, with a benefit of b and tax of t as given. We also must allow for differential crowdout effects for each type of worker, so define the crowdout of the private insurance contract for type θ as $r(\theta) = -\frac{db_p(\theta)}{db}$. The optimal utilitarian social insurance contract can be characterized as follows.

Proposition 4 When private insurance is set optimally, the marginal effect of an increase in the benefit b is given by

$$\frac{dW}{db} = -\frac{1-\overline{a}}{\overline{a}}\frac{E(au'(c_1))}{\overline{a}}\varepsilon_{1-\overline{a},b} - \frac{cov(a,u'(c_1))}{\overline{a}} \tag{1.12}$$

Proof

Optimal Private Contract. Firms, taking b and $t = \frac{1-\overline{a}}{\overline{a}}b$ as given choose $b_p(\theta)$ to maximize

$$\begin{aligned} W(\theta) =& a(\theta)u\left(M(\theta) - t - \frac{1-a(\theta)}{a(\theta)}b_p(\theta)\right) \\ &+ (1-a(\theta))u(m(\theta) + b + b_p(\theta)) - h(a(\theta)) \end{aligned} \tag{1.13}$$

Using the envelope theorem for the agents' optimal choice of a, we have that

$$\frac{dW}{db_p(\theta)}|_{b,t} = (1-a(\theta))u'(c_1(\theta))\left(\frac{u'(c_0(\theta)) - u'(c_1(\theta))}{u'(c_1(\theta))} - \frac{\varepsilon_{1-a(\theta),b_p(\theta)}|_{b,t}}{a(\theta)}\right) \tag{1.14}$$

This expression is the analog of (1.8). Each firm separately provides an optimal insurance contract for their worker subject to earning zero profits. The optimal private insurance contract, conditional on b, is therefore characterized by

$$\frac{u'(c_0(\theta)) - u'(c_1(\theta))}{u'(c_1(\theta))} = \frac{\varepsilon_{1-a(\theta),b_p(\theta)}|_{b,t}}{a(\theta)} \tag{1.15}$$

Social Insurance in the Presence of Optimized Private Insurance. The government chooses b to maximize social welfare, taking into account $b_p(\theta)$ set as in (1.15). In particular, welfare is given by

$$W = \int \left(a(\theta) u \left(M(\theta) - \frac{1-\bar{a}}{\bar{a}} b - \frac{1-a(\theta)}{a(\theta)} b_p(\theta) \right) \right.$$
$$\left. + (1-a(\theta)) u(m(\theta) + b + b_p(\theta)) - h(a(\theta)) \right) f(\theta) d\theta \tag{1.16}$$

Employing the envelope theorem for the agents' utility maximizing choice of a, we have that

$$\frac{dW}{db} = \int ((1-a(\theta)) u'(c_0(\theta))(1-r(\theta))$$
$$-a(\theta) u'(c_1(\theta)) \frac{d}{db} \left(\frac{1-\bar{a}}{\bar{a}} b + \frac{1-a(\theta)}{a(\theta)} b_p(\theta) \right) \Bigg) f(\theta) d\theta \tag{1.17}$$

Moreover,

$$\frac{d}{db} \left(\frac{1-\bar{a}}{\bar{a}} b + \frac{1-a(\theta)}{a(\theta)} b_p(\theta) \right)$$
$$= \frac{1-\bar{a}}{\bar{a}} + b \frac{d \left(\frac{1-\bar{a}}{\bar{a}} \right)}{db} - r(\theta) \frac{1-a(\theta)}{a(\theta)} + b_p(\theta) \frac{d \left(\frac{1-a(\theta)}{a(\theta)} \right)}{db}$$
$$= \frac{1-\bar{a}}{\bar{a}} + \frac{1-\bar{a}}{\bar{a}^2} \varepsilon_{1-\bar{a},b} - r(\theta) \frac{1-a(\theta)}{a(\theta)} + \frac{b_p(\theta)}{b} \frac{1-a(\theta)}{a(\theta)^2} \varepsilon_{1-a(\theta),b} \tag{1.18}$$

Plugging (1.18) into (1.17), we have that

$$\frac{dW}{db} = E((1-a)u'(c_0)(1-r)) - E(au'(c_1)) \left(\frac{1-\bar{a}}{\bar{a}} \right) \left(1 + \frac{\varepsilon_{1-\bar{a},b}}{\bar{a}} \right)$$
$$+ E((1-a)u'(c_1)r) - E \left(\frac{1-a}{a} u'(c_1) \frac{b_p}{b} \varepsilon_{1-a,b} \right) \tag{1.19}$$

Adding and subtracting the term $E((1-a)u'(c_1))$ to (1.19) we have that

$$\frac{dW}{db} = E((1-a)u'(c_0)(1-r)) - E((1-a)u'(c_1)(1-r)) + E((1-a)u'(c_1))$$
$$- E(au'(c_1)) \left(\frac{1-\bar{a}}{\bar{a}} \right) \left(1 + \frac{\varepsilon_{1-\bar{a},b}}{\bar{a}} \right) - E \left(\frac{1-a}{a} u'(c_1) \frac{b_p}{b} \varepsilon_{1-a,b} \right) \tag{1.20}$$

To impose the constraint that firms are offering optimal insurance conditional on b, we now utilize (1.15). Multiply both sides of (1.15) by $(1-a(\theta))(1-r(\theta))u'(c_1(\theta))$ and take the expectation over productivity types to obtain

$$E((1-a)(1-r)(u'(c_0)-u'(c_1))) = E\left(\frac{1-a}{a}u'(c_1)(1-r)\varepsilon_{1-a,b_p}|_{b,t}\right) \qquad (1.21)$$

Finally, we note that $\varepsilon_{1-a(\theta),b} = \varepsilon_{1-a(\theta),b_p(\theta)}|_{b,t}(1-r(\theta))\frac{b}{b_p(\theta)}$. Plug this expression into (1.21) and note that (1.21) can be substituted in for the first two terms of (1.20), which then cancels with the last term of (1.20). This yields

$$\frac{dW}{db} = E((1-a)u'(c_1)) - E(au'(c_1))\left(\frac{1-\overline{a}}{\overline{a}}\right)\left(1+\frac{\varepsilon_{1-\overline{a},b}}{\overline{a}}\right) \qquad (1.22)$$

We can further transform this expression by converting to one of covariances in the following way:

$$\begin{aligned}
\frac{dW}{db} &= E(u'(c_1)) - E(au'(c_1))\left(1+\frac{1-\overline{a}}{\overline{a}}\left(1+\frac{\varepsilon_{1-\overline{a},b}}{\overline{a}}\right)\right) \\
&= -\frac{1-\overline{a}}{\overline{a}}\frac{E(au'(c_1))}{\overline{a}}\varepsilon_{1-\overline{a},b} + E(u'(c_1)) - \frac{E(au'(c_1))}{\overline{a}} \\
&= -\frac{1-\overline{a}}{\overline{a}}\frac{E(au'(c_1))}{\overline{a}}\varepsilon_{1-\overline{a},b} - \frac{cov(a,u'(c_1))}{\overline{a}}
\end{aligned}$$

This is the expression in (1.12) which we desired. ■

It is clear that Proposition 4 provides a generalization of the result in Proposition 2. In particular, with ex-ante homogeneity of types, (1.12) reduces to (1.6) since $\overline{a} = a$, the expectation operator drops out, and the covariance between effort and marginal utility in the high state is zero. It is no longer necessarily the case, however, that it is optimal for the government to set $b = 0$ with ex-ante heterogeneous agents.

The first term in (1.12) is of the same form as in (1.6). Providing a positive level of social insurance contract reduces effort and crowds out private insurance. With homogeneity, firms provide all of the necessary insurance, and this negative effect of a marginal increase in $b > 0$ is the only effect that matters. By introducing heterogeneity, we add the second term in (1.12). We cannot sign the covariance without more structure on the nature of the shocks faced by workers. When the covariance between effort and marginal utility is positive, the marginal impact of b on welfare becomes more negative. A positive covariance is consistent with agents who consume more also exerting less effort. Thus, the ex-post distribution of production across the economy is compressed, mitigating the need for government redistribution. If we restrict $b \geq 0$, we are then at a corner solution in which the government should not provide any insurance.

If, however, the covariance between effort and marginal utility in the high state is negative, there is scope for government redistribution. It is reasonable that under some specifications, higher types will have larger consumption in the high state and also exert more effort. A negative covariance implies a positive term added to the first negative term in (1.12). Intuitively, if risks are negatively correlated with income so that the high earners have low probability of bad outcomes, there is a greater motive for redistribution through higher taxation. Although the setting is different than in previous analyses, we find that the introduction of moral hazard and an efficient private insurance market does not change the robust result that the covariance between risk and ability is a primary factor in determining the desirability of redistribution. Greater risk aversion also increases the magnitude of this covariance, whereas risk neutrality implies a zero covariance and there is hence no role for social insurance.

The expression in (1.12) can also be estimated by the econometrician to determine the marginal impact of more social insurance on welfare. In the unemployment context, although effort is not observable, the realized fraction of the population that is unemployed, $1 - \overline{a}$, is observable. In addition to estimating the elasticity of unemployment with respect to social insurance, the econometrician must estimate the covariance between effort and marginal utility of consumption. This could be estimated by considering differential unemployment rates among different classes of earners. This makes the analysis of practical and empirical relevance.

It is also instructive to compare the optimal social insurance with and without private insurance. Define s^* to be the solution to $\frac{dW}{ds} = 0$ in (1.10) and b^* to be the solution to $\frac{dW}{db} = 0$ in (1.22). They are reprinted for clarity:

$$\frac{dW}{ds} = E((1-a)u'(c_0)) - E(au'(c_1))\frac{1-\overline{a}}{\overline{a}}\left(1 + \frac{\varepsilon_{1-\overline{a},s}}{\overline{a}}\right) \tag{1.10}$$
$$\frac{dW}{db} = E((1-a)u'(c_1)) - E(au'(c_1))\frac{1-\overline{a}}{\overline{a}}\left(1 + \frac{\varepsilon_{1-\overline{a},b}}{\overline{a}}\right) \tag{1.22}$$

Despite the similarity in form of the two expressions, we cannot directly compare them since the c_0, c_1, and a are different in the two cases. Although we expect that $\varepsilon_{1-\overline{a},s} > \varepsilon_{1-\overline{a},b}$ due to crowdout, we also expect that average effort is lower with private insurance. This makes comparing the second terms in the above equations impossible without additional structure on the problem. If $\frac{dW}{ds} > \frac{dW}{db}$, we would obtain $s^* > b^*$. Such a result suggests that without firms providing insurance to individual workers, there exists a greater role for the government to provide insurance both within and across workers and the optimal benefit rate is greater than in the case in which firms are endogenously providing some insurance to individuals. If, however, $\overline{a}$ is much more responsive to s relative to b, due to a large crowdout effect or large ex-ante variance, we may have that $s^* < b^*$. We will turn to numerical simulations to better understand how the prescription for optimal policy depends on the presence of endogenously provided private insurance to individual agents.

1.4.3 Partial Private Insurance

It is reasonable to expect that not all individuals in an economy are privately insured. In this section we explore the impact of social insurance in a setting in which workers are exogenously determined to be either optimally insured by a private firm or receive no private insurance. To simplify the exposition, we assume in particular, that a share, α of each type of worker is privately insured. Given government policy, the problem for private insurers is unaffected in this new setting. The social planner's problem is now essentially a convex combination of the objective functions in Propositions 3 and 4. The only difference is that the government's budget constraint is affected by the fact that privately insured workers will exert less effort than their non-insured counterparts. We introduce the notation of superscripts of I and

N on the variables a, c_1, and c_0 to indicate the differential efforts and consumptions of insured and non-insured individuals, respectively. It is also useful to define $\overline{a^I} = \int a^I(\theta)f(\theta)d\theta$ and $\overline{a^N} = \int a^N(\theta)f(\theta)d\theta$ as the average effort among the privately insured and non-insured, respectively. Finally, we denote the elasticities of the average probability of the low event with respect to the government benefit, b, for each class of workers in the standard fashion: $\varepsilon_{1-\overline{a^I},b}$ and $\varepsilon_{1-\overline{a^N},b}$.

Proposition 5 With α share of the population privately insured, the marginal impact of a benefit b on welfare is given by

$$\begin{aligned}\frac{dW}{db} &= \alpha E((1-a^I)u'(c_1^I)) + (1-\alpha)E((1-a^N)u'(c_0^N)) \\ &\quad - \frac{\alpha E(a^I u'(c_1^I)) + (1-\alpha)E(a^N u'(c_1^N))}{\alpha\overline{a^I} + (1-\alpha)\overline{a^N}} \\ &\quad \times \Bigg[\alpha(1-\overline{a^I}) + (1-\alpha)(1-\overline{a^N}) \\ &\qquad + \frac{\alpha(1-\overline{a^I})\varepsilon_{1-\overline{a^I},b} + (1-\alpha)(1-\overline{a^N})\varepsilon_{1-\overline{a^N},b}}{\alpha\overline{a^I} + (1-\alpha)\overline{a^N}}\Bigg] \qquad (1.23)\end{aligned}$$

Proof See Appendix

The expression in (1.23) reduces to (1.10) and (1.22) when α is zero and one, respectively. This generalization characterizes the effect of α on the marginal benefit to society of increasing social insurance benefits. As discussed in the previous subsection, it is not possible to order $\frac{dW}{db}$ for $\alpha \in \{0, 1\}$. Although we cannot sign the effect of α on $\frac{dW}{db}$, the root of the expression in (1.23) maps out the change in optimal policy as α changes continuously. We will explore this issue numerically in the following section.

Despite the somewhat cumbersome expression, it follows the standard intuition for optimal taxation. The marginal benefit is captured by the gap in marginal utilities in the low and high states. In particular, the first term in (1.23) is the weighted average of the the expected marginal utility in the high state for the privately insured and the marginal utility in the low state for the uninsured weighted respectively by the probability of the low state. The marginal utility is taken at the high state as opposed to the low state for the insured due to the presence of optimal private insurance. The first part of the second term is the average expected marginal utilities in the high states for both classes of agents weighted by the probability of the high state. The gap between these two terms is a measure of the benefit to society of greater insurance. We must also adjust this second term, however, by the last term in square brackets in (1.23) to capture the behavioral effect of greater social insurance and its subsequent impact on the government's budget constraint. The form of the expression follows from the balanced budget constraint, which requires a tax of

$$t = \frac{\alpha(1-\overline{a^I}) + (1-\alpha)(1-\overline{a^N})}{\alpha\overline{a^I} + (1-\alpha)\overline{a^N}} b$$

A richer extension would allow for the distribution of types who are privately insured to be endogenously determined. In the unemployment insurance context we could think of some occupations in which employment is at will, while others offer explicit severance packages or private unemployment insurance as negotiated by a union. Optimal policy may be different if individuals can search and self-select occupations and firms in turn can determine the optimal level of insurance to attract different types of workers.[8]

1.5 Numerical Simulations

The preceding analysis provided a sharp characterization of the impact of government insurance in the presence of privately provided insurance. The tractability of the problem was due in part to the assumption that social insurance took an additive form. Thus, all agents received the same level of taxes and benefits, independent of their wages. In reality, however, we may imagine a social insurance system that relies on setting marginal tax rates. In particular, consider a social planner setting a

[8]Guerrieri, Shimer, and Wright (2009)[18] characterize equilibrium properties in a search model with ex-ante heterogeneity.

single marginal income tax rate and rebating the proceeds equally to all agents. This has the advantage of not requiring the government to observe whether an agent is in the high or low state of nature. Hence, although the unemployment context was an appropriate interpretation of the additive social insurance setting, with wage taxes we can think of agents experiencing high and low productivity shocks that do not lead to unemployment and firms offering insurance via a compressed wage structure. Such a system, however, imposes the complication of workers responding to tax policy via both income and substitution effects.

We can consider the same environments as in Section 4 with ex-ante heterogeneity and the possible provision of optimal private insurance. Without private insurance, there will exist a single social insurance system consisting of a proportional wage tax, s, and a rebate equal to $s\overline{z}$ where

$$\overline{z} \equiv \int \left(a(\theta)M(\theta) + (1 - a(\theta))m(\theta)\right) f(\theta)d\theta$$

is total output. When firms offer private insurance, they will take the same form as before with type-contingent contracts $(b_p(\theta), t_p(\theta))$ where $t_p(\theta) = \frac{1-a(\theta)}{a(\theta)} b_p(\theta)$. The government imposes a proportional wage tax of b on wages after the provision of private insurance. The rebate is then equal to $b\overline{w}$ where

$$\overline{w} \equiv \int \left(a(\theta)(M(\theta) - t_p(\theta)) + (1 - a(\theta))(m(\theta) + b_p(\theta))\right) f(\theta)d\theta$$

is aggregate earnings. Appendix A provides a setup of the social planner problem in both of these settings to obtain the marginal welfare impact of changing the wage tax. Unfortunately, the resulting expressions do not provide a clear intuition as to the how optimal tax rate is affected by the presence of private insurance. Numerical simulations, however, are instructive in this more complex environment.

1.5.1 The Impact of Private Insurance on the Level of Social Insurance

The previous theoretical and empirical literature has noted that in the presence of optimal private insurance, the optimal level of social insurance is lower (see Chetty and Saez (2009)[7], Cutler and Gruber (1996)[10], Golosov and Tsyvinski (2007)[15], and Kaplow (1991)[20]). The intuition for such a result is clear; social insurance provides insurance and redistribution across agents and with private insurance markets, part of the government's objective is achieved privately. The previous theoretical analyses have not allowed, however, for endogenous distributions of ex-post outcomes, as in this chapter. This allows for parameterizations of the model in which optimal social insurance is higher in the presence of optimal private insurance. The intuition for such a result is that private insurance compresses the wage distribution for each worker facing uncertain productivity shocks. This compression is welfare improving by smoothing consumption, but reduces the effort supplied by each agent. Lower effort choices increase the likelihood of low states of nature being realized. This exacerbates the need for redistribution to the lowest types in the economy, making the optimal social insurance level higher than if no private insurance existed.

To verify the above intuition, consider the following parameterization of the economy. Let utility, $u(\cdot)$ be given by CRRA preferences with coefficient of relative risk aversion of 3. Let the marginal disutility of effort be given by $h'(a) = -\eta \ln(1-a)$ for $\eta > 0$. This somewhat nonstandard utility representation is not significant for the results, but simply ensures an interior solution to the consumer's effort choice problem. Moreover, we will follow Mirrlees (1990)[25] in defining the shock process to be given by $M(\theta) = M\theta$ and $m(\theta) = m\theta$.[9] Let there be five uniformly distributed ex-ante types in the economy with $E(\theta) = 1$. Finally, we let $\eta = 1$, $M = 1.25$, and $m = 0.75$, but the result is robust to many alternate specifications.

Figure 1-1 plots the optimal tax rate in the presence and absence of optimal private insurance as a function of the variation in ex-ante heterogeneity (the distance between

[9]Various other specifications of the shocks yield analogous results with only slight modifications to the levels.

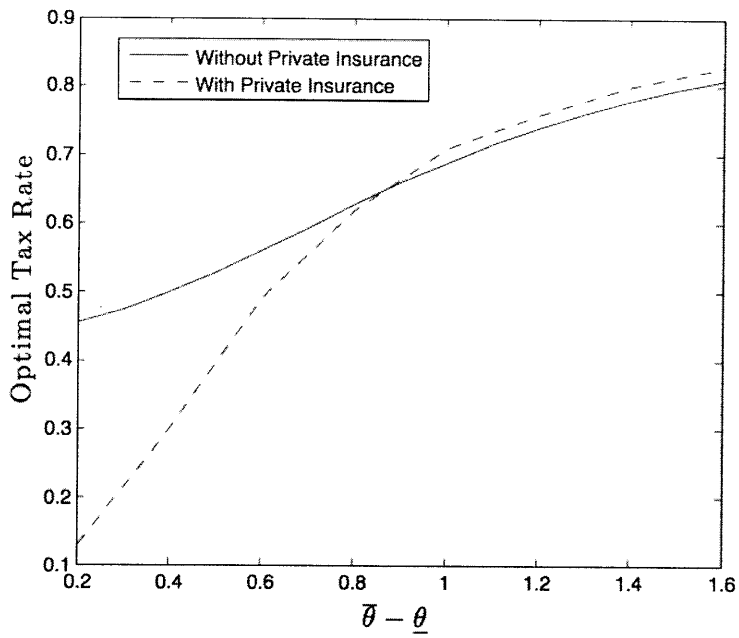

Figure 1-1: Optimal Taxation in the Presence of Private Insurance

the highest and lowest skilled types, while fixing $E(\theta) = 1$.) For small variation in the ex-ante heterogeneity, we confirm the predictions of previous analyses with an optimal level of social insurance lowered by the presence of private insurance. The optimal tax rate is increasing in the heterogeneity in the population, as the redistributive motive is greater with greater variation. As the ex-ante variation increases relative to the ex-post shocks, redistribution becomes more important relative to insurance. Since private insurance cannot redistribute between workers, the difference between optimal taxes with and without private insurance diminishes. The novel feature of this setting is that in fact, the two plots cross. It is at this point that the presence of private insurance hampers the government's redistributional goals. For large enough ex-ante heterogeneity, a social planner redistributing income with a linear wage tax simultaneously provides sufficient insurance for the uncertain outcomes. Additional private insurance further depresses effort and this greater probability of low outcomes reduces welfare more than the marginal increase in insurance. Thus, the social planner finds it optimal to redistribute more to the lowest earners with a higher marginal tax rate.

From a policy perspective, this result suggests that for economies with large variation in ability, generous insurance can exacerbate inequality, requiring higher taxation. Insurance programs for disability, for example, may induce low ability workers to choose low effort so as to receive disability benefits. Greater welfare could be potentially achieved by reducing the generosity of such benefits and lowering income tax rates.

When determining optimal policy it is therefore important to not only consider the availability of private insurance for individuals, but also the ex-ante heterogeneity of types in the economy. The welfare loss from failing to account for the existence of private insurance can be dramatic with more homogenous populations, but becomes small for more heterogeneous populations. In the next section we further investigate the role of heterogeneity on optimal policy.

1.5.2 Optimal Social Insurance without Private Insurance

To better understand the determination of the optimal marginal tax rate without private insurance, we continue with a slightly different numerical exercise. In the previous subsection we saw that optimal taxation without private insurance is increasing in the variance of the ex-ante heterogeneity. Such monotonicity, however, does not obtain for an increase in the variance of ex-post shocks. To see this, continue with the same parameterization as in the previous subsection. Consider two different levels of ex-ante heterogeneity. In one case let the variance be large, with a distance between the highest and lowest productivity types, $\overline{\theta} - \underline{\theta} = 1$ and in another case $\overline{\theta} - \underline{\theta} = 0.2$. Again, we fix a uniform distribution of types with $E(\theta) = 1$. Figure 1-2 plots the optimal marginal tax rate in these two cases without private insurance as we vary the ex-post variation in shocks. In particular, we vary m and M such that $\frac{m+M}{2} = 1$ so that shocks are symmetric around an agent's type.[10] Figure 1-2 displays the simulation.

Note that we do not have monotonicity in the ex-post variance. This is in con-

[10]This assumption does not drive the results. For example, fixing $M = 1$ and varying m from 0 to 1 yields similar results.

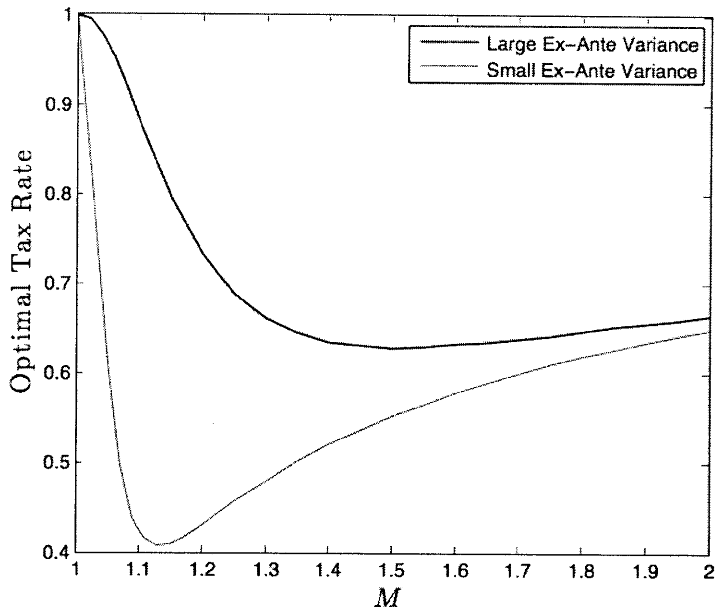

Figure 1-2: Optimal Taxation without Private Insurance

trast to Mirrlees (1990)[25] where it is shown that with two sources of heterogeneity and no private insurance, the optimal linear tax rate is increasing in both ex-ante and ex-post variation. That result does not hold in this model since the labor choice impacts the distribution of realized outcomes. To understand the intuition for such non-monotonicity, fix the ex-ante heterogeneity while introducing risk. When there is no ex-post risk and agents receive the same consumption regardless of effort, effort is optimally zero. In such a case, there is no labor distortion to taxation and the government sets a confiscatory tax rate in which incomes are equated across all individuals in the economy. Greater variation in outcomes for each type of agent increases effort, and hence a distortionary cost of taxation to the social planner. The optimal tax rate must then initially decrease. As efforts increase, the rate at which different types of agents increase effort for a given increase in risk varies. When ex-ante heterogeneity is not very large, the lowest type productivity workers exert greater effort initially relative to their richer counterparts since they need to insure themselves more against the worst outcomes. This leads to a decrease in the average output, $\overline{z}$. The lower risk for the low types due to self-insurance reduces the motive for redistribution. As the shocks increase in magnitude, higher types start increasing effort relatively more than low types who are already exerting high effort. This leads to an increase in average output in the economy. With higher output and lower risk for the highest earners due to their greater effort, the redistributive motive increases and leads to higher optimal social insurance when this outweighs the cost of the behavioral response on the government budget constraint. In addition, when the ex-ante variance is large, it requires larger negative shocks for the highest types to increase effort sufficiently high so as to induce the social planner to increase redistribution.

The preceding analysis is depicted graphically in Figures A-1 and A-2 in Appendix A and follows the same intuition we have already seen in our analytical results. Namely, when risk and income are sufficiently negatively correlated, the motive for redistribution is strengthened. Although the importance of this correlation has been established in the literature, this numerical exercise permits a better understanding of how it translates into the level of optimal taxation for various parameterizations. Note also that as the magnitude of the shocks increases, the difference between optimal social insurance at different levels of ex-ante heterogeneity decreases since the ex-post variation dominates any initial ex-ante differences.

1.5.3 The Impact of Ex-Ante Heterogeneity on Optimal Policy in the Presence of Private Insurance

It is also instructive to simulate how the optimal tax policy is affected by the presence of optimal private insurance for different levels of ex-ante heterogeneity. We continue with the same parameterization as in the previous subsection. Figure 1-3 plots the optimal marginal tax rate in the two cases of large ($\overline{\theta}-\underline{\theta}=1$) and small ($\overline{\theta}-\underline{\theta}=0.2$) ex-ante heterogeneity with and without private insurance as we vary the ex-post variation in shocks. Note that the plots of optimal taxation without private insurance are the same as in Figure 1-2. The plot also reinforces the already noted observation from Figure 1-1 that an increase in ex-ante heterogeneity leads to an increase in the optimal tax rate at all levels of ex-post shocks.

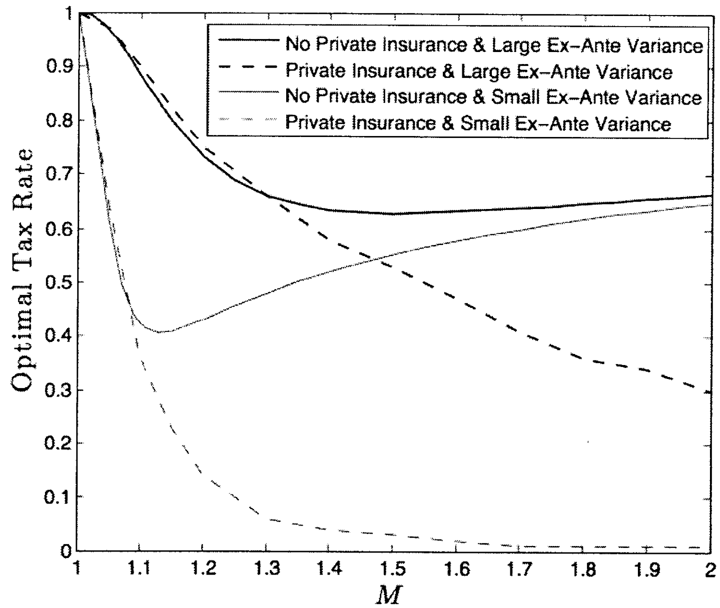

Figure 1-3: Optimal Taxation

With optimal private insurance, the optimal tax rate is in general lower than without private insurance, but not always, as discussed previously. As the ex-post variance increases, the optimal tax rate monotonically declines in both cases. Although greater variation may increase the value of redistributive taxation, the source of that variation is important. For a fixed population of workers, as ex-post shocks increase in magnitude, insurance is the primary welfare improving tool relative to redistribution. Since firms are able to insure agents against such uncertainty, the government's role as a redistributing agent is diminished and optimal taxation declines. To better understand the underlying mechanism for this result we can also examine how effort changes as the magnitude of the ex-post shocks increases. Although effort is increasing, private insurance is relatively more generous for the high type workers due to the multiplicative ex-post shock structure, leading to less effort than the low types. Since the highest types have more risk of low outcome events, the covariance between effort and marginal utility is positive, thereby weakening the case for social insurance. Appendix A provides the plots of these observations in Figures A-3 and A-4.

What is most dramatic about the numerical simulation is the gap between the optimal tax rate with and without private insurance at the two different levels of ex-ante heterogeneity. When ex-ante heterogeneity is large ($\overline{\theta} - \underline{\theta} = 1$), there is very little difference between optimal taxation levels with and without private insurance for moderate variation in ex-post shocks. Most of the variation in the economy is due to ex-ante differences, which the private market cannot smooth across. In such a case, standard optimal taxation formulas that neglect private insurance do not lead to substantial welfare losses. For $M > 1.3$ the gap becomes significant, in which case, there is a potential large welfare gain to lowering the level of social insurance. This gap, however, is clearly smaller than the analogous gap for smaller ex-ante heterogeneity. When ex-ante heterogeneity is small, optimal taxation in the presence of private insurance converges to zero quickly. The intuition for such a result lies in the fact that with ex-ante similar individuals experiencing productivity shocks, almost all of the variation in the economy is due to ex-post shocks and each agent has a very similar set of possible outcomes as their cohorts. Thus, there is very little scope for redistribution. In such an economy, there may be substantial welfare losses by failing to recognize the role of private insurance.

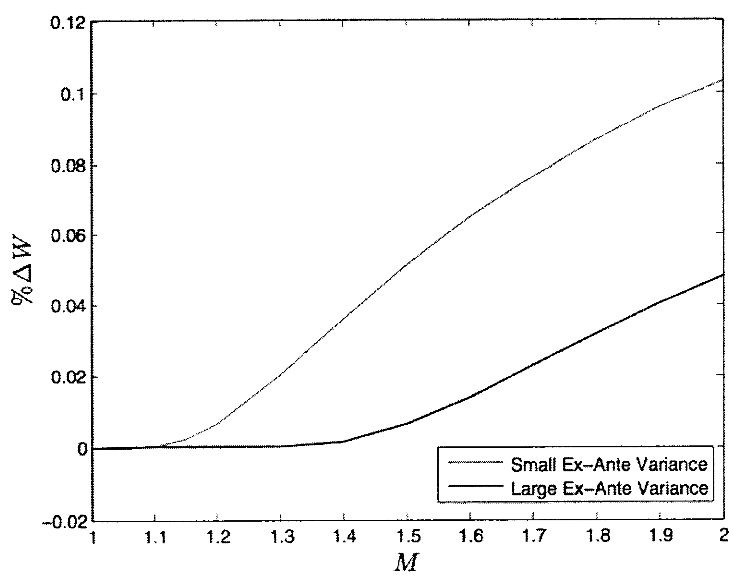

Figure 1-4: Loss In Welfare Associated with Suboptimal Social Insurance

The results of Figure 1-3 can be translated into a welfare analysis by comparing welfare under two scenarios. Assume that the economy is one in which there is a well-functioning private insurance market. First, consider the optimal social insurance policy set by a utilitarian social planner who takes into account the presence of private insurers. Second, consider a social planner setting optimal policy as if there were no private insurance market. As shown in Figure 1-3, this will in general lead to too much insurance relative to the optimum and reduce welfare. The percentage loss in welfare relative to the level of welfare in the first case is plotted in Figure 1-4. The parameterization of the economy is unchanged from that which generated Figure 1-3 and the two plots refer to the cases in which ex-ante heterogeneity is large ($\overline{\theta}-\underline{\theta}=1$) and small ($\overline{\theta}-\underline{\theta}=0.2$). As expected from the prior discussion, the welfare loss from failing to recognize endogenous private insurance markets is increasing in the variance of the ex-post shocks since private insurance is effective at smoothing consumption for a given agent across high and low states of nature. Moreover, a decrease in ex-ante heterogeneity implies that the primary source of heterogeneity is through ex-post shocks, which again can be optimally insured in the private market. A social planner ignoring such private insurance thus has a larger negative impact on welfare. In summary, welfare losses due to ignoring the private insurance market are decreasing in ex-ante heterogeneity, but increasing in ex-post heterogeneity.

1.5.4 Partial Private Insurance

In an economy in which only some workers are privately insured, it is clear that optimal tax rates should be decreasing in the fraction of the population which is insured (conditional on optimal taxes being lower in the presence of private insurance). When more individuals are able to smooth consumption with private insurance contracts, the government's tax policy serves only a redistributive role as opposed to also acting as insurance. The following simulation considers the case we have already analyzed in which ex-ante heterogeneity is large ($\overline{\theta}-\underline{\theta}=1$). We have already computed the optimal tax rates in which everyone is privately insured and no one is privately insured for various levels of heterogeneity in the ex-post shocks. In Figure 1-5, six additional intermediate cases are considered. For three of the cases, 50% of the population is privately insured and in the other three, 90% of the population is privately insured. Given the aggregate number of insured, we then consider three sub cases in which insurance is evenly distributed among the population and when insurance is perfectly positively and negatively correlated with ability. The most reasonable case is that in which the correlation is positive. High ability workers are more likely to have access to private insurance. This may be due, for example, to employers who offer high ability workers severance packages in the case of unemployment. And although adverse selection has been assumed away in this model, in actual insurance markets, we expect to see higher type individuals to have greater access to private insurance.

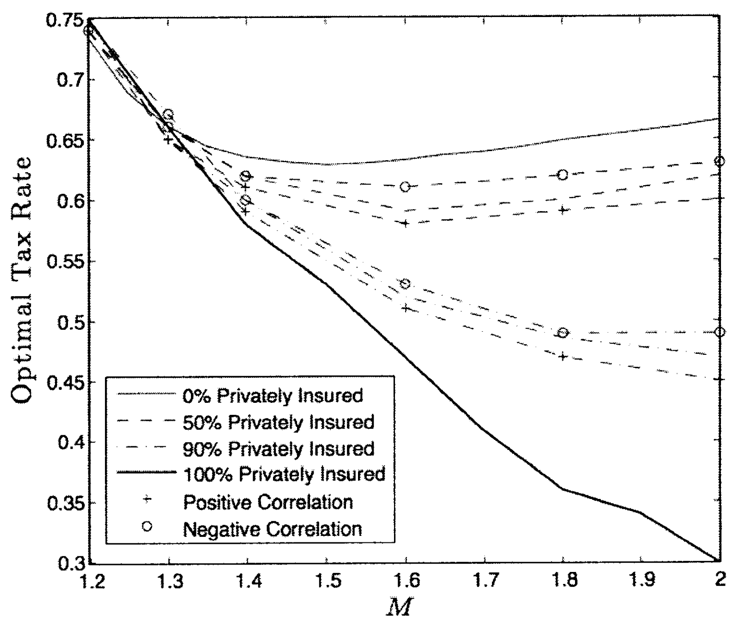

Figure 1-5: Partial Private Insurance

The results in Figure 1-5 confirm the intuition that optimal taxation is indeed decreasing in the fraction of the population with private insurance. It is interesting to note the concave relationship between the fraction of the population covered and optimal marginal tax rate. When a small fraction of the population is privately insured, an increase in private insurance coverage has a smaller impact on reducing the optimal tax rate than when most of the population is already insured. Consider the case in which $M = 1.6$. Figure 1-5 shows that randomly insuring the first half of the population leads to a decrease in the optimal tax rate of 4 percentage points, while insuring the next half leads to an additional 12 percentage point drop in the level of the optimal marginal tax rate. The intuition for this result follows from the endogeneity of the private insurance contracts. Consider private firms dropping insurance coverage for one percent of the population. When few are privately insured, the government can make up for the loss in private insurance by offering more generous insurance to all with a small increase in the marginal tax rate. This has a minimal impact on the small number of existing private insurance contracts. When most individuals in the economy are privately insured, however, providing the same additional insurance now requires a greater increase in taxes since the increase in taxation is less effective as it crowds out private insurance throughout the economy.

From a policy perspective, this result suggests two issues. First, our estimates of the welfare loss from setting taxes without accounting for endogenous private insurance markets may be overstated even with robust private insurance. Second, policies which encourage private insurance provision will lead to increasing declines in the optimal tax rate.

The numerical simulation additionally reveals that introducing a positive or negative correlation between ability and private insurance coverage while fixing the total fraction of those privately insured yields only small adjustments to the level of the optimal tax rate. We see in fact that the plots for optimal policy in the cases of correlated private insurance coverage track and bound the optimal tax rate when the correlation is zero. In particular, relative to the uniform distribution case, a positive correlation implies that more high types are insured, reducing the average effort of the high types in the population. Hence, the covariance between risk and ability is increased, weakening the redistributive role of taxation. Similarly, a negative correlation implies that the highest types are least likely to be insured and therefore exert more effort on average. This decreases the covariance between risk and ability, making more redistribution optimal.

Although the preceding observations are based on numerical simulations, they allow for a richer understanding of how the presence of private insurance markets and the degree of heterogeneity in a population can dramatically affect optimal social insurance policy. The parameterizations have been stylized to highlight important features of the model. An important next step would be an attempt to calibrate this model to the U.S. economy. The challenge inherent in any such task, however, is disentangling observed earnings into their two components of ex-ante productivity and ex-post productivity shock. As explained, the relative magnitudes of these two sources of heterogeneity is integral to the computation of optimal social insurance.

1.6 Conclusion

This chapter has characterized optimal social insurance in an economy with endogenous individual private insurance and a richer notion of heterogeneity than in previous theoretical analyses. In particular, agents differ in their ex-ante productivity and subsequently experience a shock, resulting in either a high or low state of nature, relative to to their innate productivity. Analytical results are achieved in a setting in which the government can observe the nature of the ex-post shock, as in unemployment. The chapter shows that with ex-ante homogeneous agents, the introduction of optimal private insurance contracts obviates the need for any social insurance. Private and public insurers are equally effective at providing insurance to smooth consumption for a single type of agent. However, when individuals differ ex-ante, the social planner has the extra degree of freedom of being able to redistribute across agents. Thus, there may be scope for government intervention. The optimal social insurance benefit depends on the covariance between effort and marginal utility of consumption in the high state of nature. Intuitively, if the agents who exert more effort also consume more upon the realization of the high state, the social planner can improve a utilitarian welfare objective function by redistributing to those with lower earnings.

To supplement the analytical results, the chapter also provides numerical simulations of the economy in a more general and less tractable setting in which the governing authority needs only to observe realized earnings and not the type of ex-post shock in order to implement its policy. The natural justification for this setting is one in which the government imposes a linear income tax on a wage distribution that has been compressed by firms offering insurance to implement optimal effort provision. The simulations highlight a number of interesting features. The first is that although in many cases the presence of private insurance reduces optimal social insurance, this may not always hold. Since private insurance reduces effort, making low state events more likely, the redistributive motive may increase in the presence of private insurance when ex-ante heterogeneity is sufficiently large. Also, the welfare loss from optimizing social insurance, but ignoring private insurance markets is greatest when agents are ex-ante similar, but face large shocks, since in that case, the primary welfare improving tool is consumption smoothing within workers and not redistribution across workers. Thus, the presence of private insurance is effective at providing all of the insurance for the economy.

There exist interesting directions to expand this line of research theoretically. A simplifying assumption of the analysis in this chapter is that private insurers are able to observe innate productivity. Such an observation eliminates the adverse selection issue. The introduction of adverse selection interacting with the moral hazard problem is an important step in better understanding the endogenous response of private insurance contracts to social insurance provision.[11] In addition, if productivity is imperfectly observable or there are administrative costs, private insurance may not be provided optimally as assumed throughout this chapter.[12] Non-optimized private insurance may also come in the form of informal insurance mechanisms available to individuals, such as the ability to borrow from friends or rely on spousal labor income, as documented by Cullen and Gruber (2000)[9]. Implicit throughout this analysis and much of the related literature on optimal insurance in the presence of private insurance is that private insurers earn zero profits. One could consider instead a situation in

[11]Boadway et. al. (2006)[4] consider a case with adverse selection and ex-post moral hazard (conditional on being in the low state, agents can spend money to improve their welfare in that state) as opposed to the ex-ante moral hazard problem in this chapter (effort is chosen before uncertainty is realized to impact the probability of the high state).

[12]Chetty and Saez (2009)[7] investigate non-optimized private insurance in a simpler setting with only one source of heterogeneity.

which firms are not risk neutral and there exists a risk sharing contract between worker and firm. In such a setting, the firm becomes an agent in the economy and one could investigate if social insurance is as effective at redistributing across workers or simply affects the extent of risk sharing between firm and worker.

This chapter is an important step in better understanding the factors which affect the determination of optimal social insurance when private insurance contracts and individual behavior respond endogenously. An important, but challenging direction for future empirical work is to calibrate the model with taxation as social insurance to a real economy. Doing so would require a disentangling of the observed variation in realized earnings into their unobserved ex-ante variation in skills and ex-post shocks to productivity. The implication for policy is that we may in fact be able to improve welfare by reducing taxes that serve as a redistributive mechanism.

Chapter 2

Efficiency Wages with Heterogeneous Agents

2.1 Introduction

The presence of persistent involuntary unemployment in many labor markets requires an explanation as to why these markets do not clear. Noncompetitive market structures, government policies, and search frictions each provide possible reasons as to why wages fail to clear labor markets. The seminal paper by Shapio and Stiglitz (1984) [28] (henceforth SS) presents a model of efficiency wages in which higher wages create an unemployment pool, which serves as a discipline device to prevent workers from shirking on the job. This chapter extends that basic model by allowing for a more plausible model with worker heterogeneity. The assumption of identical workers provides a clear and intuitive framework for understanding a market failure induced by moral hazard. As is evident by casual observation, however, workers vary along many dimensions. Individuals differ with respect to skill, experience, work history, drive, education, and a host of other demographic factors.

In order to provide a simple, but plausible notion of heterogeneity in the population, workers are assumed to be equally productive while exerting effort on the job, but to differ with respect to their disutility of labor effort. Thus, while some workers may enjoy their work, others may find labor sufficiently intolerable as to prefer shirking on the job. Firms have the option of then offering contracts which induce effort from all, or from only some workers. This chapter provides conditions for the existence of each of these types of potential pure strategy symmetric Nash equilibria (PSSNE) and characterizes the contracts offered therein. The analysis shows that it is possible, depending on the degree of heterogeneity among workers, for there to exist a unique or no PSSNE.

When workers are sufficiently similar, there will exist a unique PSSNE in which all workers exert effort, as in the original SS framework. The maximum allowable extent of heterogeneity to generate such an equilibrium is relative to the distribution of worker types in the economy. When the economy consists of more individuals who are less likely to shirk, allowing shirking from a lazy, but small population becomes less costly. Thus, worker heterogeneity must decrease to guarantee a unique PSSNE without shirking. However, as heterogeneity increases, the existence of a PSSNE depends on the equilibrium accession rate at which the unemployed move into employment. When the accession rate is low, firms allowing shirking from some workers will have a higher proportion of effort-exerting workers than for a higher accession rate. Thus, for low accession rates, a shirking equilibrium will exist, while for higher accession rates, a no-shirking equilibrium will exist.

In addition, I extend the analysis in the second half of the chapter to allow for a labor hours decision in conjunction with the effort provision decision. This is an additional departure from the standard efficiency wage model, going back to Yellen (1984) [31]. The classical labor model tends to fix effort and allow labor hours to vary, whereas most shirking moral hazard models fix labor hours, with effort as the choice variable. A more general approach, which I pursue here, is for both effort and labor hours to be variable. This allows for a more realistic characterization of the multidimensional problem faced by workers. Such an extension introduces the joint problem of adverse selection to the existing moral hazard problem. Firms must not only offer wages to induce workers to find it optimal not to shirk on the job, but firms must also screen workers based on their unobservable disutility of labor effort. I find that such a generalization, however, does not change the analysis in which labor hours are fixed. In particular, I am able to show that there is no PSSNE in which different types of workers work different hours. This is due to the fact that offering a separating contract is only optimal for a firm wishing to induce effort from all workers and in addition, the workers are sufficiently heterogeneous. If workers are too different, however, then the firm would prefer to allow the most labor-averse workers to shirk, breaking the proposed equilibrium. This model thus predicts that differences in labor hours across similar occupations cannot be found within a firm, but may only exist in asymmetric equilibria.[1]

The modeling assumptions of this chapter follow many of the simplifying assumptions used in SS so as to isolate the implications of introducing worker heterogeneity. This chapter is not the first, however, to introduce such heterogeneity. Strand (1987) [29] introduces worker heterogeneity in the extreme case that one type of worker will always shirk. This setting allows Strand to study how firms might screen workers based on past employment and firing histories. Albrecht and Vroman (1998) [2] (henceforth AV) likewise take up the task of extending SS to a setting with workers who differ with respect to their disutility of effort. Their main result is that there cannot exist a pure strategy symmetric Nash equilibrium in wage offers. Thus, a wage dispersion equilibrium must prevail. Their analysis, however, rests crucially on the assumption of a continuum of worker types. The analysis of this chapter illustrates that such a result fails to hold with a discrete number of types in the population. In addition, I provide the extension of firms offering a menu of multidimensional labor contracts to workers, who then self-select.[2] Although such an extension does not generate a new PSSNE, it answers the question left open at the end of AV as to how such issues of adverse selection would impact the efficiency wage setting. Therefore, the determination of the equilibrium relies heavily on the specification of the heterogeneity in the population and not on the scope of the firms' ability of offer multidimensional contracts.

[1]One could also interpret an asymmetric pure strategy Nash equilibrium as a symmetric mixed strategy Nash equilibrium, but the former interpretation seems more natural.

[2]Following SS and AV, I do not allow for an alternative contract, such as workers posting performance bonds, so as to again, isolate the implications of heterogeneity on the efficiency wage setting.

The outline of the chapter is as follows. The next section sets up the framework of the model. Section 3 characterizes the economy with fixed labor hours, Section 4 introduces variable labor hours for potential deviant firms, Section 5 considers the general case with flexible labor hours for all firms, and Section 6 concludes. Lengthy proofs are relegated to Appendix B for clarity of exposition.

2.2 The Model

The model follows the framework of Shapiro and Stiglitz (1984) [28] and its extension by Albrecht and Vroman (1998) [2]. The setting is in continuous time with infinitely-lived individuals discounting the future at rate $r > 0$.

2.2.1 Workers

There is a continuum of workers in the economy, which is fixed with measure N. There exist two types of employees, with disutility of labor effort for the high and low types given by $\theta_h, \theta_l \in \mathbb{R}_{++}$, respectively. We assume that $\theta_l > \theta_h$ so that by "high" type we mean those workers with a lower disutility of effort. A fraction $\pi \in (0, 1)$ of the workers are high type individuals. Instantaneous utility for consumers is linear and given by $y - \theta L e$, where y is income, $L \in [h, H]$ is labor hours supplied, and e is effort per hour worked.[3] Additionally, we assume that the effort choice is discrete and in particular, $e \in \{0, 1\}$.[4]

Workers in this economy, as in SS, can either be employed or unemployed. While employed, workers earn an hourly wage w and thus income, $y = wL$. Shirkers are detected and fired by firms at an exogenous rate, $d > 0$. Workers can also suffer an

[3]We can interpret the lower and upper bounds on work hours as corresponding to part-time and full-time work, respectively, or analogously as full-time and over-time work, for example. Since we introduce a linear technology in the next section, firms will in in general find corner solutions to be optimal. Thus, although we allow for interior solutions, h or H will always be in any set of hours offered in a profit maximizing labor contract.

[4]Pisauro (1991) [26] explains in his model without a labor supply choice that with utility linear in consumption, it is without loss of generality to assume that effort takes on two extreme values, as it will never be optimal for the firm to induce partial effort by the workers. This follows the SS assumptions of linear utility and discrete effort choice.

exogenous separation from a firm, which occurs at rate $s > 0$. The Poisson parameters d and s are fixed scalars.

If unemployed, $e = 0$ and we assume that consumption while unemployed is normalized to zero as well since unemployment benefits policy is not the focus of this model. Unemployed individuals receive job offers at a rate $a > 0$. This Poisson offer arrival rate is endogenously determined in a steady state equilibrium, but is exogenous from the point of view of an individual. Moreover, matching to firms is strictly random. In particular, a certain number of vacancies per unit time will be created in equilibrium due to exogenous separations and possible firings that result from detection of shirking. In order to maintain their steady-state workforces, firms will recruit new workers from the unemployment population. In continuous time, the probability that an unemployed individual will be recruited is equal to the measure of created vacancies relative to the measure of unemployed, due to the Poisson assumption. The assumption of random matching gives that this probability is the same for all workers. Thus, length of unemployment spell, past employment history, and disutility of labor do not impact the probability of being matched to a particular firm. Firms cannot observe past employment/firing histories of workers, so no stigma is attached to previously shirking workers.[5]

An individual matched to a firm selects a contract consisting of a wage/hour pair, (w, L) from the menu of contracts, C, offered by the firm. The worker, upon being matched to a firm, can decide to reject the match and return to the unemployment pool, or conditional on accepting the job, decide whether or not to shirk. Note that the firm is not permitted to fire a worker conditional on the contract they choose. The following value functions characterize the worker's problem:

$$V_N(w, L; \theta) = \frac{(w - \theta) L}{r + s} + \frac{s}{r + s} U(\theta) \tag{2.1}$$

[5]See Strand (1987) [29] for an extension of SS with heterogeneous workers in which firms can screen workers depending on the extent to which workers who were previously fired can find a new job relative to those who quit.

$$V_S(w, L; \theta) = \frac{wL}{r+s+d} + \frac{s+d}{r+s+d} U(\theta) \tag{2.2}$$

$$U(\theta) = \frac{a}{r+a} E \max\{ \max_{(\xi,\delta)\in C} \max_{i\in\{N,S\}} \{V_i(\xi, \delta; \theta)\}, U(\theta)\} \tag{2.3}$$

where the expectation in (2.3) is taken with respect the distribution of contract menu offers across firms.

The first value function, V_N, is the value to a worker of type θ accepting a contract (w, L) at a firm and exerting effort on the job. The second value function, V_S, indicates the value to a worker who decides to shirk on a job while receveing the contract (w, L). The final value function measures the value to an unemployed worker. The formulation of the value functions follows from standard dynamic programming and allows for the possibility that firms offer a distribution of contracts. This setup is analogous to AV, differing only in the allowance of contracts to consist of labor hours. Firms must take into account the decision process of workers when faced with labor contract offers.

2.2.2 Firms

There is a continuum of endogenous measure M of identical firms that sell a homogeneous numeraire good in a competitive market. In steady state equilibrium with a constant unemployment rate, firms earn zero profit and thus M will adjust accordingly. Firms face a pool of unemployed workers, to whom they offer an efficiency wage contract, which consists of an hourly wage, w, and number of hours, L, demanded by the firm. Labor hours spent on the job are observable by the firm, such as through a time stamp. However, firms can neither observe a worker's disutility of labor effort nor their effort level, except via the monitoring technology. In addition, although firms can choose the size of their workforce, the relative composition of high to low types in the workforce is determined in equilibrium.

Firms have identical production functions, given by

$$f(\overline{L_h} + \overline{L_l})$$

where f is continuously differentiable, strictly increasing, and strictly concave. We assume that the only costs faced by the firms are the variable labor costs and a fixed flow cost c. The upper bars over the L_i are used to indicate the total labor supply of each type of worker employed by the firm. In particular, $\overline{L_i} = e_i n_i L_i$ where n_i is the number of workers of type i in the firm.

The specification of the firm's profit function, along with the heterogeneity with respect to the disutility of labor effort yields the following result which helps to simplify the analysis.

Proposition 1 If a menu of contracts is rejected by some workers and accepted by others, then any worker who would accept a contract with the firm would always shirk.

Proof This proof is analogous to the argument in AV's Proposition 2. Suppose a firm, m, offers a menu of contracts, $C_m = \{(w, L)\}$ that is rejected by a worker of type θ_i. This implies that $U(\theta_i) > \max\{V_S(w, L; \theta_i) : (w, L) \in C_m\}$. This is equivalent to $rU(\theta_i) > \max\{wL : (w, L) \in C_m\}$. Some workers must accept a contract in equilibrium. Thus, workers of type $\theta_j \neq \theta_i$ select a contract and accept employment. Let us denote the optimal contract selected by worker of type θ_j to be $(w_j, L_j) \in C_m$. Thus, $w_j L_j \geq rU(\theta_j)$, otherwise the worker would reject the offer. Hence, $U(\theta_i) > U(\theta_j)$ and since $U(\cdot)$ is a decreasing function (see AV for proof), $\theta_i < \theta_j$. It follows from $rU(\theta_i) > w_j L_j$ that $V_S(w_j, L_j; \theta_i) > V_N(w_j, L_j; \theta_i)$. We also know from differentiating the value functions that $\frac{\partial V_N(w,L;\theta)}{\partial \theta} \leq \frac{\partial V_S(w,L;\theta)}{\partial \theta}$. Therefore, $V_S(w_j, L_j; \theta_j) > V_N(w_j, L_j; \theta_j)$. Hence, any worker who accepts would shirk. ■

The preceding proposition implies that no contracts are offered in equilibrium in which some workers choose to reject the offer, since all other workers would shirk.

It cannot be in a firm's interest to offer contracts in which all of its workers shirk. Therefore, there is no job search in the sense that upon being recruited by a firm, no worker rejects in favor of waiting for another job offer. This model thus allows us to isolate the role of the moral hazard problem in an efficiency wage model, as distinct from other search considerations, to generate unemployment.

This simplification of the problem allows us to write the value function for the unemployed as

$$U(\theta) = \frac{a}{r+a} E\left(\max_{(\xi,\delta)\in C} \max_{i\in\{N,S\}} V_i(\xi,\delta;\theta) \right) \tag{2.3$'$}$$

Firms still have the discretion to choose contracts in which some workers may or may not find it optimal to shirk while employed.

With the workers and firms specified, we can now turn to characterizing the existence and type of pure strategy symmetric Nash equilibria that will prevail in this labor market with both adverse selection and moral hazard constraints.

2.3 Equilibrium with Fixed Labor Hours

We begin our analysis of this economy by making the simplifying assumption in this section that labor hours are fixed. In particular, we will assume that $L = 1$ for all workers and the problem of the firm becomes one of simply offering one dimensional wage contracts, as in AV. We will look for pure strategy symmetric Nash equilibria, which do not exist in the AV framework. The analysis in AV, however, relies crucially on the assumption of a continuum of worker types, as opposed to a discrete number of types as in this chapter.

Before proceeding, it is helpful to introduce notation for the "no-shirk condition" for workers. Workers will not shirk when $V_N(w;\theta) \geq V_S(w;\theta)$, which is equivalent to $w \geq rU(\theta) + \frac{r+s+d}{d}\theta$.[6] We can define the critical value $\theta_N(w)$ by

$$w = rU(\theta_N(w)) + \frac{r+s+d}{d}\theta_N(w) \tag{2.4}$$

[6]We drop the labor argument in this section since it is fixed.

From Proposition 1 of AV we know that any worker of type $\theta \leq \theta_N(w)$ will put forth effort at a firm offering a wage w, while any worker of type $\theta > \theta_N(w)$ will shirk at this wage. We will now turn to considering two potential PSSNE: one in which all workers exert effort at all firms, and one in which only the high type workers exert effort.

2.3.1 No Shirking Equilibrium

Consider the problem of a firm in an economy in which all other firms set a wage, $\tilde{w}$, such that $V_N(\tilde{w};\theta) \geq V_S(\tilde{w};\theta) \quad \forall\theta$. In this case, we can solve for the unemployment value function:

$$U(\theta) = \frac{a}{r(r+a+s)}(\tilde{w} - \theta) \quad \forall\theta \tag{2.5}$$

Combining this with the expression in (2.4) gives

$$w = \frac{a}{r+a+s}(\tilde{w} - \theta_N(w)) + \frac{r+s+d}{d}\theta_N(w) \tag{2.6}$$

which implicitly defines $\theta_N(w)$ as an increasing function of w.

A no shirking PSSNE will prevail if all firms indeed find it optimal to induce effort from all workers by offering the same wage, without deviating to a different wage offer in which only the high types exert effort. We will determine the profits in these two cases in turn.

Symmetric Best Response

A firm seeking to induce effort from all of its employees must solve the following problem:

$$\max_{w,n} \quad f(n) - wn - c \tag{2.7}$$
$$s.t. \quad \theta_h, \theta_l \leq \theta_N(w)$$

where n is the number of employees in the firm. The optimal wage offer will be defined by $\theta_N(w) = \theta_l$, or $w = \frac{a}{r+a+s}(\tilde{w} - \theta_l) + \frac{r+s+d}{d}\theta_l$. In a symmetric equilibrium, $w = \tilde{w}$. Solving this for the equilibrium wage, w^*, we obtain

$$w^* = \frac{r+a+s+d}{d}\theta_l \tag{2.8}$$

The comparative statics have the same interpretation as in SS. A better monitoring technology, as expressed with a higher d, makes shirking more costly for workers and hence lowers firm costs. Increases in r, a, and s all serve to lower the value of being employed relative to being unemployed, thus making shirking less costly for workers and subsequently increasing costs for firms wishing to induce effort from workers.

Potential Deviation

It is necessary to also check the conditions under which a firm would not prefer to offer a lower wage in which only the high types exert effort. A deviation to a lower wage will apply to all workers at the firm, not simply the new hires. Thus, at the instant of deviation, wage costs will fall and the low type workers will prefer to shirk at the new wage (we will verify below that no worker prefers to quit from a deviating firm). Over time, however, the high type workers, the only ones providing productive labor, will constitute a greater fraction of the deviating firm's labor force, since the low types will be exiting the firm at a greater rate. Thus, as long as the new steady state profit level for a deviating firm is no greater than the profit from not deviating, we will obtain the labor contracts as a Nash equilibrium. Such a condition is not necessary, however, if for instance the firm were to have a high discount rate and the transition process to the new steady state took sufficiently long. Following AV, however, we will implicitly assume a near zero discount rate for firms so that the comparison between steady state profit levels will be both a necessary and sufficient condition for a Nash equilibrium.

Hence, given that all other firms are offering w^* as in the previous section and that workers will not reject the new wage offer, a potential deviating firm's profit in the new steady state is given by the following value function

$$\max_{w,n} \quad f(\psi n) - wn - c \tag{2.9}$$
$$s.t. \quad \theta_h \leq \theta_N(w) < \theta_l$$

where ψ is the fraction of the firm's workers who are high type. The low type workers will be fired by the firm at a higher rate than they are hired from the unemployment pool, leading to the firm increasing its fraction of high type workers relative to non-deviating firms. High type workers will be leaving the deviating firm at rate s, whereas low type workers will be leaving at rate $s+d$. The deviating firm is small and therefore does not affect the distribution of types in the unemployment population, where the fraction of unemployed who are high types is still π. Thus, the probability that a vacancy is filled by a high type worker in the deviating firm is given by π. Since the flow of workers into the deviating firm will be greater than the flow into non-deviating firms, we can understand the filling of vacancies as being directed by an auctioneer. A deviating firm wishing to maintain a given firm size, n, will thus adjust to a new distribution of types determined by

$$n\psi s = \pi n(s\psi + (s+d)(1-\psi))$$

$$n(1-\psi)(s+d) = (1-\pi)n(s\psi + (s+d)(1-\psi))$$

Solving, yields that

$$\psi = \pi\frac{s+d}{s+\pi d} \geq \pi \tag{2.10}$$

It becomes clear that the closer that ψ is to one, the greater the incentive to deviate. Deviation allows for a lower wage, but with less production. The degree to which production falls will depend on how far ψ deviates from π.

As before, the optimal wage will solve $\theta_h = \theta_N(w)$, or

$$\begin{aligned} w &= \frac{a}{r+a+s}(w^* - \theta_h) + \frac{r+s+d}{d}\theta_h \\ &= \frac{r+a+s+d}{d}\left[\frac{(r+s)\theta_h + a\theta_l}{r+a+s}\right] \end{aligned} \tag{2.11}$$

which is clearly a lower wage than that in (2.8). Note that for θ_h and θ_l very similar, the wage difference by a deviating firm is likewise very small. For greater heterogeneity between types, the weighted average of types in the brackets in (2.11) falls, inducing a lower wage offer by a deviating firm. With a larger gap in types, it is more likely that a firm may prefer to reduce wage costs significantly by allowing shirking from some workers. Moreover, the difference between w^* in (2.8) and the deviating wage offer in (2.11) is greater for faster outflows from employment and slower inflows into employment. Intuitively, if workers leave employment quickly and remain in unemployment for some time, then deviating firms are able to lower the wage even more to still induce effort from the highest types who have worse prospects after employment.

Although we know from Proposition 1 that no offer will be rejected in equilibrium, we must verify that the above deviation does not in fact lead to workers choosing to reject the proposed wage offer. It is straightforward to establish that this is indeed the case, as the following inequalities illustrate.

$$\begin{aligned} & V_N(w;\theta_h) \geq U(\theta_h) \\ \Leftrightarrow\ & \frac{w - \theta_h + sU(\theta_h)}{r+s} \geq U(\theta_h) \\ \Leftrightarrow\ & w - \theta_h \geq rU(\theta_h) = \frac{a}{r+a+s}(w^* - \theta_h) \text{ by (2.5)} \\ \Leftrightarrow\ & \frac{a}{r+a+s}(w^* - \theta_h) + \frac{r+s}{d}\theta_h \geq \frac{a}{r+a+s}(w^* - \theta_h) \text{ by (2.11)} \end{aligned}$$

which is true.

In addition, low types will prefer shirking at the deviating firm to rejecting the offer.

$$
\begin{aligned}
& V_S(w;\theta_l) \geq U(\theta_l) \\
\Leftrightarrow\ & \frac{w+(s+d)U(\theta_l)}{r+s+d} \geq U(\theta_l) \\
& \Leftrightarrow w \geq rU(\theta_l) = \frac{a}{r+a+s}(w^* - \theta_l) \text{ by (2.5)} \\
\Leftrightarrow\ & \frac{a}{r+a+s}(w^* - \theta_h) + \frac{r+s+d}{d}\theta_h \geq \frac{a}{r+a+s}(w^* - \theta_l) \text{ by (2.11)} \\
\Leftrightarrow\ & \frac{a}{r+a+s}(\theta_l - \theta_h) + \frac{r+s+d}{d}\theta_h \geq 0
\end{aligned}
$$

which is again true.

Conditions for Symmetric Nash Equilibrium

For there to be a pure strategy symmetric Nash equilibrium with all workers exerting effort, the preceding deviation cannot not profitable. In particular, a necessary and sufficient condition for the proposed equilibrium is

$$
\begin{aligned}
& \max_n f(n) - \frac{r+a+s+d}{d}\theta_l n - c \\
& \quad \geq \max_n f\left(\pi\frac{s+d}{s+\pi d}n\right) - \frac{r+a+s+d}{d}\left[\frac{(r+s)\theta_h + a\theta_l}{r+a+s}\right] n - c
\end{aligned}
$$

Note that firms have already minimized the cost of effective labor in each case and are now choosing the size of their firm. Thus, this inequality is satisfied if and only if the cost of effective labor is weakly greater in the proposed deviation. In particular, an equivalent necessary and sufficient condition is

$$
\frac{\frac{r+a+s+d}{d}\left[\frac{(r+s)\theta_h + a\theta_l}{r+a+s}\right]}{\pi\frac{s+d}{s+\pi d}} \geq \frac{r+a+s+d}{d}\theta_l \tag{2.12}
$$

Simplifying (2.12) yields the following equivalent condition for an equilibrium:

$$
\frac{\theta_h}{\theta_l} \geq \frac{\pi(s+d)}{s+\pi d} - \frac{as(1-\pi)}{(s+\pi d)(r+s)} \tag{2.13}
$$

Moreover, since the right hand side of (2.13) is linearly decreasing in a, a sufficient condition for the equilibrium is

$$\frac{\theta_h}{\theta_l} \geq \frac{\pi(s+d)}{s+\pi d} \tag{2.14}$$

Intuitively, types must be sufficiently similar relative to the fraction of high types that will prevail in a deviating firm so that no firm can gain by ignoring effort from those workers with strong disutility of labor and still retaining a large productive workforce. This is in contrast to the result in AV in which there is no symmetric Nash equilibrium when there is a continuum of worker types. Note that even in the limit of types becoming very similar in this model, we only strengthen the likelihood of this symmetric equilibrium obtaining.

Implicit in this analysis is that firm profits are zero at the equilibrium. Since equilibrium profits are decreasing in a, as long as the fixed cost is not so large as to induce all firms to exit the industry, there will be a unique accession rate which yields zero profits. In steady state equilibrium the flow in and out of unemployment must be equal, i.e., $Mns = (N - Mn)a$ where n is again the size of each firm's workforce and M is the measure of firms operating. This condition therefore determines M.

We now turn to investigating the existence of another type of PSSNE.

2.3.2 Shirking Equilibrium

Consider the problem of a firm in an economy in which all other firms set a wage, $\tilde{w}$, such that $V_N(\tilde{w};\theta_h) \geq V_S(\tilde{w};\theta_h)$ and $V_N(\tilde{w};\theta_l) \leq V_S(\tilde{w};\theta_l)$. Solving for the unemployment value function:

$$U(\theta_h) = \frac{a}{r(r+a+s)}(\tilde{w} - \theta_h) \tag{2.15}$$
$$U(\theta_l) = \frac{a}{r(r+a+s+d)}\tilde{w} \tag{2.16}$$

The value functions here differ for the two types since the low type workers are now shirking and future employment involves the risk of being detected and fired at rate

d.

We now follow analogous steps as in the previous section to determine conditions under which there will indeed be no profitable deviation by a firm to induce effort from all workers.

Symmetric Best Response

A firm choosing to induce effort from only the high types will solve:

$$\begin{aligned} \max_{w,n} \quad & f(\hat{\psi}n) - wn - c \\ s.t. \quad & \theta_h \leq \theta_N(w) < \theta_l \end{aligned} \tag{2.17}$$

Plugging (2.15) into (2.4) we have that the optimal wage that solves $\theta_N(w) = \theta_h$ is given by $w = \frac{a}{r+a+s}(\tilde{w} - \theta_h) + \frac{r+s+d}{d}\theta_h$. In a symmetric equilibrium, $w = \tilde{w}$ and solving we obtain

$$w^* = \frac{r+a+s+d}{d}\theta_h \tag{2.18}$$

Note that the wage in (2.18) is of the same form as the no-shirking equilibrium wage in (2.8), except here we have θ_h instead of θ_l. The relevant constraint in this case is to induce effort from only the high type, not both high and low, so we have a lower wage offer. Otherwise, the comparative statics analysis of the model parameterization on the wage is analogous to before.

We must also determine the steady state value of $\hat{\psi}$, the fraction of high types in a firm. Using the steps in Lehr (2010) [22],

$$\pi \leq \hat{\psi} = \frac{\pi(a+s+d)}{a+s+\pi d} \leq \psi \tag{2.19}$$

There are more high types in a firm in this setting than when all workers are exerting effort in all firms since the low types are exiting the firm at a faster rate than the high types. However, the fraction of high types in a firm that allows shirking is less than the case in which all other firms are inducing effort since when all other firms are also permitting shirking, the unemployment pool is populated with more shirkers, increasing the likelihood of hiring a low type.

Potential Deviation

A firm that decides to deviate to offering a higher wage in which all of its employees exert effort when all other firms are offering w^* as in (2.18) will solve:

$$\max_{w,n} f(n) - wn - c \tag{2.20}$$
$$s.t. \quad \theta_h, \theta_l \leq \theta_N(w)$$

where we initially assume that workers will accept the wage offer. Note that the deviation here is instantaneous: higher wages induce all workers to exert effort and since non-shirking workers do not differ with respect to productivity, the share of the deviating firm's work force providing labor is always 100%. Although the deviating firm will shed low type workers at a slower rate than all other firms and hence have a relatively greater fraction of low type workers, there is no transition path to the new steady profit level. Hence, we do not need to appeal to a near zero discount rate to generate the sufficient condition of a lower new steady state profit level as a necessary condition as well.

The optimal wage will solve $\theta_N(w) = \theta_l$. Using (2.4) and (2.16) we have that

$$w = \frac{a\theta_h + (r + s + d)\theta_l}{d} \tag{2.21}$$

which is a higher wage than the candidate equilibrium wage in (2.18). Again the firm must weigh the benefit of higher effort from all against the cost of higher wages for all of its employees. The gap between the deviating wage in (2.21) and the wage in (2.18) depends on the extent of heterogeneity. If types are similar, a small wage increase is sufficient to induce effort from the low types. If, however, the low type has a much greater aversion to labor effort than the high type, a larger wage increase must be offered. It is also clear that with a higher wage offer in this potential deviation, no worker would choose to reject such an offer. Thus, it was legitimate to assume that no workers would prefer unemployment to staying employed by a firm making such a deviation.

Conditions for Symmetric Nash Equilibrium

Compiling the preceding expressions for wages and steady state shares of productive workers, a necessary and sufficient condition for a pure strategy symmetric Nash equilibrium in which all firms induce effort from only the high type workers is that

$$\max_n f\left(\frac{\pi(a+s+d)}{a+s+\pi d}n\right) - \frac{r+a+s+d}{d}\theta_h n - c \geq \max_n f(n) - \frac{a\theta_h + (r+s+d)\theta_l}{d}n - c$$

or equivalently,

$$\frac{a\theta_h + (r+s+d)\theta_l}{d} \geq \frac{(r+a+s+d)(a+s+\pi d)}{d\pi(a+s+d)}\theta_h \tag{2.22}$$

which can be expressed also as

$$\frac{\theta_l}{\theta_h} \geq \frac{a+s+d\pi}{\pi(a+s+d)} + \frac{a(a+s)(1-\pi)}{\pi(a+s+d)(r+s+d)} \tag{2.23}$$

The right hand side of (2.23) is increasing without bound in a. Thus, for high equilibrium accession rates, a firm allowing shirking has fewer high types and in order to prevent such a firm from offering a wage that induces effort from all workers, the low type must have a sufficiently high disutility of labor relative to the high type to make deviating to a non-shirking wage offer for all too costly.

Note that the equilibrium a will again be uniquely determined by the point at which profits of firms permitting shirking is zero. The equilibrium measure of firms, M, will be given by the steady state condition of equal flows in and out of unemployment: $Mn\hat{\psi}s = (N\pi - Mn\hat{\psi})a$, where n is the number of employees in a single firm.

2.3.3 Characterization of Symmetric Nash Equilibria

The preceding analysis shows the conditions under which there exist pure strategy symmetric Nash equilibria. We can summarize the results in the following proposition.

Proposition 2

a. If $\frac{\theta_h}{\theta_l} \geq \psi = \frac{\pi(s+d)}{s+\pi d}$, then there exists a unique pure strategy symmetric Nash equilibrium in which all workers exert effort.

b. If $\frac{\theta_h}{\theta_l} < \psi$, then there exists at most one pure strategy symmetric Nash equilibrium and in particular

   i. For $c > 0$ sufficiently small, all workers exert effort in the unique PSSNE

   ii. For $c > 0$ sufficiently large, all workers shirk in the unique PSSNE

   iii. For intermediate values of $c > 0$, there does not exist a PSSNE

Proof See Appendix B.

Proposition 2 shows us that we can retain the original motivation of efficiency wages to induce effort from all workers even with heterogeneous workers. Such an equilibrium is guaranteed when the heterogeneity among the types is not too varied. Otherwise, with very different types, it would be in the interest of firms to offer lower wages in which the most labor-averse workers shirk. In addition to the degree of variation of the types, characterization of the PSSNE also depends on the relative rates of firing relative to exogenous separations. When separations occur at a faster rate than firings conditional on shirking (i.e., s large and d small), ψ is reduced and there is a greater range of heterogeneity in which we are guaranteed the unique symmetric Nash equilibrium with no shirking. Intuitively, in such a situation, a potential deviating firm cannot significantly gain by realizing a large relative influx of high type workers since all workers, both shirkers and not, leave the firm at a similar rate. Similarly, for more low types (π small), there exists a greater cost to allowing shirking from said low types, making a wider range of heterogeneity consistent with a unique PSSNE in which no one shirks.

When types are sufficiently different to make $\frac{\theta_h}{\theta_l} < \psi$, characterization of the equilibrium becomes more nuanced. For low fixed costs, profits are higher and more firms will enter the industry, increasing the equilibrium accession rate, a. With a higher accession rate, the deviation to offering a lower wage and allowing shirking becomes less profitable, as evidenced in (2.13). Thus, no-shirking is optimal in equilibrium. For higher fixed costs, profits are diminished and firms exit, yielding a lower accession rate. In the case in which firms allow low types to shirk in equilibrium, a smaller a implies a greater proportion of productive workers in a non-deviating firm, making shirking optimal in equilibrium. Moreover, Proposition 2 establishes that there can be no case in which both types of equilibria exist for a given parameterization of the model. There does exist a range of c, however, in which there is no pure strategy symmetric Nash equilibrium in wage offers. Note that the analogous non-existence result from AV relied on there being a continuum of types in the economy, independent of the fixed cost. Here, with only two types, we see that non-existence can also obtain, but only for sufficiently heterogeneous types and a particular range of costs.

In the following section we investigate how the option to offer two dimensional labor contracts impacts the equilibrium analysis.

2.4 Introduction of Variable Labor Hours

The preceding analysis provided a clear characterization of pure strategy symmetric Nash equilibria in an efficiency wage model with two types of workers. Following the previous literature on efficiency wages, the labor supply choice consisted only of choosing effort, not hours. However, it is reasonable to consider the possibility that firms may offer part time and full time labor contracts to separate workers based on their differing disutility of labor effort.

In this section we check to see whether a firm can profitably deviate to offering a two-part labor contract, breaking either of the PSSNE of the previous section with fixed hours. The labor hours choice is one of selecting $L_i \in [h, H]$ for $i = l, h$. Normalize $H = 1$ so that we can understand the wage offers in the previous section as incomes for full time work. It will be shown, however, that offering a part time job with h hours will never break the equilibrium. The intuition for this result stems from the observation that firms will only prefer to offer two-part labor contracts when workers are sufficiently heterogeneous. In the case of the no shirking PSSNE, when workers are sufficiently different, a firm will prefer to deviate to allowing shirking before inducing effort from all with a two-part contract. Similarly, the worker heterogeneity required to maintain a two-part contract deviation from a potential shirking PSSNE will yield shirking as the best option. Thus, for a PSSNE in which firms are offer the same hours to all workers, allowing for variable labor hours will not change the equilibrium. Both cases are analyzed in turn.

2.4.1 No Shirking Equilibrium

Consider the no shirking PSSNE with fixed hours consisting of wage offers given by (2.8). In addition, such an equilibrium holds if and only if condition (2.13) is satisfied. With the introduction of variable labor hours, a firm can deviate to a two-part contract. Note that a firm deviating to allow for shirking from the lowest type workers will not offer a two-part contract. The firm will simply offer a wage high enough for the high type worker to exert effort and all workers will accept the same contract. Such a deviating wage is in fact the wage in (2.11). The only relevance for a two-part contract is the case in which a deviating firm still wishes to induce effort from all workers. The firm may find it beneficial to lower its costs by separating workers with an hours choice, while still inducing effort from all.

Since workers are exerting effort at all other firms, from (2.5) and (2.8) we obtain the value of unemployment as

$$U(\theta_i) = \frac{a}{r(r+a+s)} \left(\frac{r+a+s+d}{d} \theta_l - \theta_i \right) \tag{2.24}$$

A deviating firm wishing to induce effort from all must solve, assuming the deviating contracts will not be rejected by the employees in favor of unemployment, the following.

$$\max_{L_l, L_h \in [h,H], n, y_h, y_l} f(n(\pi L_h + (1-\pi)L_l) - n(\pi y_h - (1-\pi)y_l) - c \tag{2.25}$$

$$s.t. \quad V_N(\frac{y_h}{L_h}, L_h; \theta_h) \geq V_N(\frac{y_l}{L_l}, L_l; \theta_h) \tag{2.26}$$

$$V_N(\frac{y_h}{L_h}, L_h; \theta_h) \geq V_S(\frac{y_h}{L_h}, L_h; \theta_h) \tag{2.27}$$

$$V_N(\frac{y_h}{L_h}, L_h; \theta_h) \geq V_S(\frac{y_l}{L_l}, L_l; \theta_h) \tag{2.28}$$

$$V_N(\frac{y_l}{L_l}, L_l; \theta_l) \geq V_N(\frac{y_h}{L_h}, L_h; \theta_l) \tag{2.29}$$

$$V_N(\frac{y_l}{L_l}, L_l; \theta_l) \geq V_S(\frac{y_h}{L_h}, L_h; \theta_l) \tag{2.30}$$

$$V_N(\frac{y_l}{L_l}, L_l; \theta_l) \geq V_S(\frac{y_l}{L_l}, L_l; \theta_l) \tag{2.31}$$

Note that in a steady state equilibrium all workers leave firms at the same rate since there is no shirking, causing the share of each firms' high workers to be equal to the population share, π. The constraints (2.26) - (2.31) are the incentive compatibility and "no-shirking constraints" for the workers. Constraints (2.27) and (2.31) are the no shirking constraints. Conditional on no shirking, firms seeking to induce effort from all agents must set contracts such that each type self-selects the contract intended for their type (constraints (2.26) and (2.29)). There is of course the possibility of a "double deviation" in which workers choose the contract not intended for their type and subsequently shirk.[7] To prevent such deviations, we impose constraints (2.28) and (2.30). Employing the expression for the value of unemployment in (2.24), we can rewrite constraints (2.26) - (2.31) as follows:

[7]This notion of a joint deviation is present in other settings of adverse selection with private information, as explored in the literature on the implementation of allocations in dynamic economies. See Albanesi Sleet (2006) [1], Golosov Tsyvinski (2006) [16], and Kocherlakota (2005) [21].

$$y_h - \theta_h L_h \geq y_l - \theta_h L_l \tag{2.26'}$$

$$y_h \geq \frac{r+s+d}{d}\theta_h L_h + \frac{a}{r+a+s}\left(\frac{r+a+s+d}{d}\theta_l - \theta_h\right) \tag{2.27'}$$

$$(r+s+d)(y_h - \theta_h L_h) \geq (r+s)y_l + \frac{ad}{r+a+s}\left(\frac{r+a+s+d}{d}\theta_l - \theta_h\right) \tag{2.28'}$$

$$y_l - \theta_l L_l \geq y_h - \theta_l L_h \tag{2.29'}$$

$$(r+s+d)(y_l - \theta_l L_l) \geq (r+s)y_h + a\theta_l \tag{2.30'}$$

$$y_l \geq \frac{r+s+d}{d}\theta_l L_l + \frac{a}{d}\theta_l \tag{2.31'}$$

The relevant constraints in this cumbersome constraint set are constraints (2.26) and (2.30), along with a monotonicity constraint that $L_h \geq L_l$. The high type workers' best deviation is to take the contract intended for the low type worker, whereas the low type worker prefers a "double deviation" by taking the high type contract while simultaneously shirking. By simplifying the constraint set we find that the solution will consist of offering the maximum number of hours to the high type, whereas the number of hours offered to the low type will depend on the relative cost of paying for an hour of labor from the low type worker. If the cost is sufficiently high, the firm will prefer to offer two dimensional contracts with $L_l < H$. Otherwise, the firm will offer all workers the same contract with full time hours. To prevent a deviation to allowing for shirking, (2.13) holds, which implies that workers are sufficiently similar. When workers are similar, however, the relative cost of inducing effort from low type labor is reduced. It is reduced just enough, in fact, that the firm will prefer to offer the same hours to everyone and we do not disrupt the equilibrium. Clearly, since the best deviation which induces effort is the equilibrium contract, it will not be rejected by workers, as assumed. This analysis is summarized in the following proposition.

Proposition 3 If a no shirking pure strategy symmetric Nash equilibrium without variable labor hours exists, then no firm will prefer to deviate to offering variable hour labor contracts if given the option.

Proof See Appendix B.

2.4.2 Shirking Equilibrium

In this section we similarly check that the introduction of variable labor hours as a possible deviation from a shirking PSSNE with fixed hours will never affect the equilibrium. The equilibrium of interest is the one in which wages are given by (2.18) and condition (2.23) is satisfied. Note that offering the equilibrium labor contract is the optimal contract in the case of allowing shirking. Thus, the only potential deviation to consider is for a firm to offer contracts which induce effort from all of its workers. In doing so, the firm now has the option to offer a two-part contract to separate workers via labor hours.

Since high type workers are exerting effort and low type workers are shirking at all other firms, we can use the unemployment value functions in (2.15) and (2.16), along with the wage (2.18) to obtain

$$U(\theta_h) = U(\theta_l) = \frac{a}{rd}\theta_h \tag{2.32}$$

The wage is such that workers face the same value of unemployment despite high types exerting effort and low types shirking.

A firm, taking the wage in (2.18) and the value of unemployment in (2.32) as given, must then solve the following profit maximization problem.

$$\max_{L_l, L_h \in [h,H], y_l, y_h, n} f(n(\widetilde{\psi} L_h + (1-\widetilde{\psi})L_l)) - n(\widetilde{\psi} y_h + (1-\widetilde{\psi})y_l) - c \tag{2.33}$$

$$s.t. (2.26) - (2.31)$$

where it is assumed and will be verified that indeed workers choose to accept contracts from such a deviation. Using (2.32), constraints (2.26)-(2.31) are equivalent to:

$$y_h - \theta_h L_h \geq y_l - \theta_h L_l \tag{2.26''}$$

$$y_h \geq \frac{r+s+d}{d}\theta_h L_h + \frac{a}{d}\theta_h \tag{2.27''}$$

$$(r+s+d)(y_h - \theta_h L_h) \geq (r+s)y_l + a\theta_h \tag{2.28''}$$

$$y_l - \theta_l L_l \geq y_h - \theta_l L_h \tag{2.29''}$$

$$(r+s+d)(y_l - \theta_l L_l) \geq (r+s)y_h + a\theta_h \tag{2.30''}$$

$$y_l \geq \frac{r+s+d}{d}\theta_l L_l + \frac{a}{d}\theta_h \tag{2.31''}$$

The fraction of high types in a deviating firm inducing effort from all when all other firms allow shirking from the low types is given by $\widetilde{\psi}$. This will be less than $\hat{\psi}$, since low types will be retained in the firm and thus compose a greater share of the workforce in a firm inducing effort from all. We can determine the steady state value of $\widetilde{\psi}$ by observing that since all of a deviating firms' workers leave at the same rate, the $\widetilde{\psi}$ must equal the proportion of high types among the unemployed. From the work in Lehr (2010) [22], it follows that

$$\widetilde{\psi} = \frac{s}{a+s}\frac{\pi}{\frac{s}{a+s}\pi + \frac{s+d}{a+s+d}(1-\pi)} = \frac{s\pi(a+s+d)}{s(a+s+d)+ad(1-\pi)} < \hat{\psi}$$

Lemma 1 The solution to (2.33), given the wage contract in (2.18) offered at all other firms, is characterized as follows:

1. $L_h = 1$

2. $L_l \in [h, 1]$

3. $y_h = \frac{r+a+s+d}{d}\theta_h + \frac{r+s+d}{d}(\theta_l - \theta_h)L_l$

4. $y_l = \frac{a}{d}\theta_h + \frac{r+s+d}{d}\theta_l L_l + \frac{r+s}{d}\theta_h(1-L_l)$

where $L_l = h$ only if

$$\frac{\widetilde{\psi}}{1-\widetilde{\psi}} \geq \frac{\theta_h \left(\frac{r+a+s}{d} + \widetilde{\psi} \right)}{\frac{r+s}{d}(\theta_l - \theta_h) + \theta_l - \widetilde{\psi}\theta_h} \tag{2.34}$$

$L_l = 1$ only if

$$\frac{\widetilde{\psi}}{1-\widetilde{\psi}} \leq \frac{\theta_h \left(\frac{r+a+s}{d} + \widetilde{\psi} \right)}{\frac{r+s}{d}(\theta_l - \theta_h) + \theta_l - \widetilde{\psi}\theta_h} \tag{2.35}$$

and $L_l \in (h, 1)$ only if

$$\frac{\widetilde{\psi}}{1-\widetilde{\psi}} = \frac{\theta_h \left(\frac{r+a+s}{d} + \widetilde{\psi} \right)}{\frac{r+s}{d}(\theta_l - \theta_h) + \theta_l - \widetilde{\psi}\theta_h} \tag{2.36}$$

Proof See Appendix B.

Moreover, the deviant contracts in the preceding analysis will not be rejected by any worker in favor of unemployment. Thus, ignoring the potential for quits in the deviant firm's profit maximization problem was legitimate. This result is established in the following lemma.

Lemma 2 For any case of optimal contracts $\{(y_i, L_i)\}_{i=l,h}$ in Lemma 1, we have that no worker will reject an employment offer from such a deviant firm . In particular, $V_N\left(\frac{y_i}{L_i}, L_i; \theta_h\right) \geq U(\theta_i)$ for $i = l, h$.

Proof See Appendix B.

Note that when a single uniform contract is the optimal deviation, then we have the same analysis as in the case with fixed labor hours. When a two-part contract is the optimal deviation, we find that firms offer a menu of contracts in which the high types work longer hours and earn higher incomes ($y_h - y_l = \theta_h(1 - L_l) \geq 0$), but earn lower hourly wages in order to prevent low type workers from shirking and choosing the labor contract intended for the high types:

$$w_l - w_h = \frac{r+s+d}{d}(\theta_l - \theta_h)(1 - L_l) + \frac{r+a+s}{d}\theta_h\left(\frac{1}{L_l} - 1\right) \geq 0$$

Thus, hours are allowing firms to separate workers who differ with respect to their distaste for labor effort. In addition, relative to the equilibrium income, a higher income must be paid to the high type worker to maintain effort and prevent a deviation to the higher wage paid to the low type worker. The income paid to the low type may be higher or lower than the income received when shirking was allowed. If workers are similar and the best deviation is one of pooling, then income for the low type will increase, maintaining the same wage across workers. If, however, the firm's best deviation is to offer separating contracts, then it is possible for the firm to generate higher wages for the low type by lowering both income and hours.

From the preceding analysis, we see that there exist two types of deviations for a firm when all other firms are offering the wage contract in (2.18) in which only the high types exert effort. As long as there is sufficient heterogeneity between the two types, we will obtain a symmetric shirking equilibrium. In fact, the constraint on guaranteeing such an equilibrium is the same constraint as in the case with no hours choice (constraint (2.23)). Thus, even though firms have the option to deviate to a separating contract, it will never be a profitable deviation when constraint (2.23) is satisfied. Intuitively, separating is only optimal when the types are sufficiently different, but when the types are different, it is preferable to simply allow the low types to shirk. The following proposition summarizes these observations.

Proposition 4 If a shirking pure strategy symmetric Nash equilibrium without variable labor hours exists, then no firm will prefer to deviate to offering variable hour labor contracts if given the option.

Proof See Appendix B.

2.5 Non-Existence of Separating Equilibrium

In the analysis of the previous section we allowed firms in a pure strategy symmetric Nash equilibrium with fixed labor hours an extra degree of freedom in terms of the space of possible deviations. In particular, by introducing a range of possible labor hours, firms gained the option to offer two-part labor contracts in order to separate the two types of workers in the economy. Propositions 3 and 4 show, however, that deviating to a two-part contract is never better than offering the equilibrium wage with the same hours for all workers. As discussed, the intuition for this result stems from the observation that firms will only find two-part contracts preferable when they are seeking to induce effort from all workers and workers are sufficiently heterogeneous, but it is precisely in this case that firms will prefer to allow shirking from the low type workers.

In this section, we consider the general case in which all firms can offer variable labor hours within their menu of contracts. The potential value of setting different labor hours is that firms can separate workers by the length of the workday. Again, let the feasible set of labor hours be given by the interval, $[h, H]$, where the natural interpretation is that of part-time work consisting of h hours and full-time work consisting of H hours. We will see, however, that there will not exist a pure strategy symmetric Nash equilibrium in which each worker type selects a distinct labor contract. The intuition from the previous section still carries through to the general case here. For a PSSNE in which firms offer two-part contracts that separate workers, workers must be sufficiently heterogeneous for firms to find such separation preferable to offering a single contract that similarly induces effort from all. However, the degree of heterogeneity needed for such a separating contract implies that firms would in fact prefer to allow the low type workers, who have a much greater disutility of labor effort, to shirk, thus destroying the proposed equilibrium. The steps of this argument follow in this section, with the details of the main proof relegated to Appendix B.

Although the analysis of the preceding sections restricted the equilibrium contracts to take the form of single wage offers to all workers, this is without loss of generality in some cases. In particular, for a shirking PSSNE, it will never be optimal to separate workers. Firms will simply offer the cheapest wage to induce effort from the high type workers and allow the low type workers to take the same contract and shirk. Thus, the equilibrium contract offer in such a PSSNE must be

$$y_h^* = y_l^* = \frac{r+a+s+d}{d}\theta_h H \qquad L_h^* = L_l^* = H \tag{2.37}$$

which is identical to the wage offer in (2.18) with H normalized to unity. And from Proposition 4 we know that when constraint (2.23) is satisfied, this is indeed a PSSNE and the availability of labor hours does not break the equilibrium.

In the case of a no-shirking PSSNE, firms have the option of inducing effort from all by offering a single contract, as in the previous section, or by offering a two-part contract with which to separate workers. It is the two-part separating contract that is the new type of PSSNE which is possible in the setting with full variable hours for all firms. We will characterize these contracts and their possible deviations in the next two lemmas.

Consider the case in which all other firms offer an identical set of contracts, $\{(\widetilde{w}_h, \widetilde{L}_h), (\widetilde{w}_l, \widetilde{L}_l)\}$, such that workers of type θ_i prefer to exert effort under contract $(\widetilde{w}_i, \widetilde{L}_i)$ than choose any other contract, shirk or both: $V_N(\widetilde{w}_i, \widetilde{L}_i; \theta_i) \geq V_k(\widetilde{w}_j, \widetilde{L}_j; \theta_i)$ $\forall i, j \in \{l, h\}$ and $k \in \{N, S\}$. The value of unemployment is therefore solved as

$$U(\theta_i) = \frac{a}{r(r+a+s)}(\widetilde{y}_i - \theta_i \widetilde{L}_i) \tag{2.38}$$

where $\widetilde{y}_i \equiv \widetilde{w}_i \widetilde{L}_i$.

A firm, facing an economy with such contracts, must set the profit maximizing contracts subject to inducing effort from all or allowing shirking from the low type workers. A Nash equilibrium requires the former to be more profitable. The profit maximization problem is precisely problem (2.25) again, with the constraints given by (2.26)-(2.31). Employing the value of unemployment in (2.38), an equivalent set of constraints is given by:

$$y_h - \theta_h L_h \geq y_l - \theta_h L_l \tag{2.26'''}$$

$$y_h \geq \frac{r+s+d}{d}\theta_h L_h + \frac{a}{r+a+s}(\widetilde{y}_h - \theta_h \widetilde{L}_h) \tag{2.27'''}$$

$$(r+s+d)(y_h - \theta_h L_h) \geq (r+s)y_l + \frac{ad}{r+a+s}(\widetilde{y}_h - \theta_h \widetilde{L}_h) \tag{2.28'''}$$

$$y_l - \theta_l L_l \geq y_h - \theta_l L_h \tag{2.29'''}$$

$$(r+s+d)(y_l - \theta_l L_l) \geq (r+s)y_h + \frac{ad}{r+a+s}(\widetilde{y}_l - \theta_l \widetilde{L}_l) \tag{2.30'''}$$

$$y_l \geq \frac{r+s+d}{d}\theta_l L_l + \frac{a}{r+a+s}(\widetilde{y}_l - \theta_l \widetilde{L}_l) \tag{2.31'''}$$

We seek to find a solution to the problem in (2.25) that is a symmetric best response to the contracts offered by the other firms. The following lemma provides a characterization of such a solution.

Lemma 3 A symmetric best response for a firm solving (2.25), given (2.38), is characterized as follows:

1. $L_h^* = H$
2. $L_l^* \in [h, H]$
3. $y_h^* = \frac{r+a+s+d}{d}\left(\theta_h H + (\theta_l - \theta_h)L_l^*\right)$
4. $y_l^* = \frac{r+a+s+d}{d}\theta_l L_l^* + \frac{r+a+s}{d}\theta_h(H - L_l^*)$

where $L_l^* = h$ only if

$$\frac{\pi}{1-\pi} \geq \frac{\theta_h\left(\frac{r+s}{d} + \pi\right) + \frac{a}{dH}\left(\theta_h H + (\theta_l - \theta_h)h\right)}{\frac{r+s}{d}(\theta_l - \theta_h) + \theta_l - \pi\theta_h} \tag{2.39}$$

$L_l^* = H$ only if

$$\frac{\pi}{1-\pi} \leq \frac{\theta_h\left(\frac{r+s}{d} + \pi\right) + \frac{a}{dH}\left(\theta_h H + (\theta_l - \theta_h)H\right)}{\frac{r+s}{d}(\theta_l - \theta_h) + \theta_l - \pi\theta_h} \tag{2.40}$$

and $L_l^* \in (h, H)$ only if

$$\frac{\pi}{1-\pi} = \frac{\theta_h \left(\frac{r+s}{d} + \pi\right) + \frac{a}{dH}\left(\theta_h H + (\theta_l - \theta_h) L_l^*\right)}{\frac{r+s}{d}(\theta_l - \theta_h) + \theta_l - \pi\theta_h} \tag{2.41}$$

Proof See Appendix B.

For the case in which the symmetric best response is one in which all firms offer a single contract to all workers, the contract simplifies to

$$y_h^* = y_l^* = \frac{r+a+s+d}{d}\theta_l H \qquad L_h^* = L_l^* = H \tag{2.42}$$

which is the same equilibrium contract as in (2.8) with H normalized to unity. And from Proposition 3 we know that if this is an equilibrium in the setting with fixed hours, a deviation employing a two-part contract can do no better. Thus, it is without loss of generality to restrict attention to fixed hour contracts when firms are inducing effort from all workers via a single contract offer.

Of interest is the whether we can establish a two-part contract which separates workers who all find it preferable to exert effort on the job. As with the contracts in Lemma 1, separation here entails higher incomes and hours for the high types, but higher wages for the low types. Note that in general we will expect a separating symmetric best response to be one in which low type workers are offered the minimum number of hours. In order to obtain an interior value for the labor hours of the low type in a symmetric best response, we must have (2.41) satisfied for some $L_l^* \in (h, H)$ and accession rate, a. However, in equilibrium the accession rate a will be pinned down by by the zero profit condition. Thus, it is unlikely that the equilibrium a will also satisfy (2.41). One could construct a parameterization of the model, however, such that this is the case.

Given the symmetric separating contracts offered as in Lemma 3, we must check whether it would be profitable for a firm to instead offer labor contracts which permitted shirking, yielding lower output and also lower costs. Assuming for the moment that workers will not reject the contracts in the deviant firm, such a deviating firm must solve:

$$\max_{L_l, L_h \in [h,H], n, y_h, y_l} \quad f(n\psi L_h) - n(\psi y_h - (1-\psi) y_l) - c \tag{2.43}$$

$$s.t. \quad (2.26) - (2.28)$$

$$V_S\left(\frac{y_l}{L_l}, L_l; \theta_l\right) \geq V_N\left(\frac{y_h}{L_h}, L_h; \theta_l\right) \tag{2.44}$$

$$V_S\left(\frac{y_l}{L_l}, L_l; \theta_l\right) \geq V_S\left(\frac{y_h}{L_h}, L_h; \theta_l\right) \tag{2.45}$$

$$V_S\left(\frac{y_l}{L_l}, L_l; \theta_l\right) \geq V_N\left(\frac{y_l}{L_l}, L_l; \theta_l\right) \tag{2.46}$$

where again $\psi = \pi\frac{s+d}{s+\pi d}$ is the fraction of high types in a firm that allows low types to shirk when all other firms in the economy induce effort from all workers. Constraints (2.26)-(2.28) are the same as in problem (2.25) in which the firm must induce effort from a high type. Constraints (2.44)-(2.46) impose that the low type worker must choose the contract intended for her and find it preferable to shirk rather than choose any other contract while shirking or working. Of these last three constraints, the likely candidate for a binding constraint is (2.45) which acknowledges that a low type shirker could also choose the contract intended for the high type and shirk. The set of constraints (2.26)-(2.28) are equivalent to (2.26‴)-(2.28‴) since the firm is still inducing effort from the high types, but (2.44) - (2.46) are equivalent to, using the value of unemployment in (2.38),

$$y_l \geq \frac{r+s+d}{r+s}(y_h - \theta_l L_h) - \frac{ad}{(r+s)(r+a+s)}(y_l^* - \theta_l L_l^*) \tag{2.44$'$}$$

$$y_l \geq y_h \tag{2.45$'$}$$

$$y_l \leq \frac{r+s+d}{d}\theta_l L_l + \frac{a}{r+a+s}(y_l^* - \theta_l L_l^*) \tag{2.46$'$}$$

Lemma 4 The solution to (2.43), given the separating symmetric contracts offered by all other firms as specified in Lemma 3, is characterized by a pooling contract in which

1. $L_h = L_l = H$

2. $y_h = y_l = \frac{r+a+s+d}{d}\theta_h H + \frac{a}{d}\frac{r+a+s+d}{r+a+s}(\theta_l - \theta_h)L_l^*$

Proof See Appendix B.

We can compare this contract with the equilibrium labor contract in which low type workers shirk at all firms. In that case, $y_h = y_l = \frac{r+a+s+d}{d}\theta_h H$ as in (2.37). Here, since low type workers at all other firms earn higher wages to induce effort, income payments to high type workers must also be higher at those firms. Thus, to still induce effort from high types, a deviant firm must offer a higher income to those workers. The extent of the markup in income depends on the relative heterogeneity in the population, with greater heterogeneity requiring a higher markup. The intuition again follows from the observation that with greater heterogeneity, it costs more for firms paying equilibrium wages to induce effort from low types and simultaneously keep the high types from choosing the contract intended for the low type. Thus, high types earn more at firms paying equilibrium wages when heterogeneity is large, requiring a deviating firm wishing to still maintain effort from the high types to pay more.

The preceding analysis was contingent on workers not electing to reject the offers of the deviating firm. We must verify that it is indeed the case that workers prefer the contract offer in Lemma 4 to returning to the unemployment pool. This is indeed the case and follows from the specification of the contracts in Lemmas 3 and 4.

Lemma 5 For the contracts $\{(y_i, L_i)\}_{i=l,h}$ in Lemma 4, we have that no worker will reject an employment offer from such a deviant firm . In particular, $V_N\left(\frac{y_h}{L_h}, L_h; \theta_h\right) \geq U(\theta_h)$ and $V_S\left(\frac{y_l}{L_l}, L_l; \theta_l\right) \geq U(\theta_l)$.

Proof See Appendix B.

In order to maintain the menu of contracts $\{(y_i^*, L_i^*)\}_{i=l,h}$ as a pure strategy symmetric Nash equilibrium, it must be the case that no firm prefers to deviate and offer the pooling contract of Lemma 4 in which only the high type workers exert effort. Note that there are two opposing constraints in this case. The types of workers must be sufficiently different so as to make it preferable for firms to offer a separating contract, but not so different as to induce firms to prefer to deviate by not paying a higher wage to the low type workers. These two constraints cannot simultaneously hold in this model and we have the non-existence of a pure strategy separating symmetric Nash equilibrium.

Proposition 5 There does not exist a pure strategy separating symmetric Nash equilibrium.

Proof See Appendix B.

The collection of Propositions 3, 4 and 5 provide a complete refutation of the need for variable labor hours in this efficiency wage model with heterogeneous workers. Together, these propositions show that the characterization of the pure strategy symmetric Nash equilibria in this model is invariant to the set of feasible labor hours. Regardless of whether labor hours for the same occupation must be equal to some fixed level, such as a forty hour work week, or can take varying values from some interval for different types of workers, there will never be a PSSNE in which workers of varying disutilities of labor effort choose to work different hours. Thus, despite the scope for screening workers in an economy with moral hazard, firms will never find it optimal to separate workers. This underscores the importance of the moral hazard problem relative to the adverse selection issue in an efficiency wage model.

2.6 Conclusion

This chapter has shown that one can construct an economy with heterogeneous workers in which firms offer the same efficiency wage contracts that induce all workers to exert effort on the job. In order to guarantee this result, we simply need to ensure that worker heterogeneity is not sufficiently varied. Thus, the only job separations in this economy are due to exogenous firings and not due to shirking. This provides an extension to the SS efficiency wage model with heterogeneous worker types in which the threat of being fired for shirking creates a non-clearing labor market.

In addititon, the chapter shows that there are parameterizations of the economy in which there may exist another equilibrium with shirking from the most labor averse workers. When types of workers are sufficiently different, this equilibrium can obtain. These results are in contrast to the AV result, in which there is no pure strategy symmetric Nash equilibrium. The difference in results is due only to the modeling of the distribution of worker types in the economy. This chapter assumes a discrete number of types, whereas AV assume a smooth distribution of types with no mass points. Which assumption is a better approximation of reality is left to the reader, but this chapter starkly illustrates the importance of this modeling assumption in delivering very different conclusions.

The interaction of adverse selection with the moral hazard constraints in an efficiency wage model remained open questions in the analyses of both SS and AV. This chapter, in its search for pure strategy symmetric Nash equilibrium, shows that under fairly general modeling assumptions, the inclusion of variable labor hours and multidimensional contracting does not affect the equilibrium set. This result may not hold with more general production technologies, but in the space of this model, we learn that we should not expect to see differences in labor hours among equally productive individuals within a given firm. The presence of part time workers in the real world economy suggests that such workers must differ from their full-time counterparts along another dimension than disutility of labor effort.

There exist many directions with which to extend this research. It will be interesting to characterize potential dispersion equilibrium in this model with and without variable labor hours. In addition, the perfect substitutes production technology, albeit a reasonable benchmark for comparison with the previous literature, may not best capture the labor contracting decisions of a firm with complementary labor inputs. A challenging but worthwhile extension would be to take seriously the issue that there is not random matching and in fact, firms can observe workers' employment histories and screen them based on this information.

In future research, it would also be interesting to investigate how optimal income taxation operates in such a setting with heterogeneous agents and involuntary unemployment. Since the canonical Mirrlees (1971) [24] optimal taxation framework assumes that all agents are simply paid a wage equal to their marginal productivity, an efficiency wage setting provides an opportunity to understand taxation when wages are the solution of a moral hazard problem. Previous research has investigated linear taxes in an efficiency wage model. Yellen (1984) [31], Johnson and Layard (1986) [19], and Pisauro (1991) [26] each provide various differing assumptions in an efficiency wage setting to investigate the comparative statics of ad valorem (wage tax) and specific (employment tax) taxes on the wages and unemployment rates. Chang (1995) [5] extends the Pisauro (1991) [26] model to include two production sectors and a labor choice decision by workers, in addition to a continuous effort supply decision. It is in this setting that Chang (1995) [5] attempts to characterize the optimal ad valorem and consumption taxes in this economy. The results suggest that consumption taxation may be more efficient than labor taxes. All of these analyses, however, restrict attention to homogeneous workers and linear taxes. Thus, this chapter serves as a potential starting point for understanding nonlinear income taxation in a more complex economic environment.

Chapter 3

Correction: Albrecht and Vroman's Nonexistence Proof of Symmetric Nash Equilbrium with Efficiency Wages

3.1 Introduction

Albrecht and Vroman (1998) [2] build an equilibrium model of efficiency wages with worker heterogeneity. It is assumed that there exists a continuum of worker types in the economy, with workers differing with respect to their disutility of labor effort. The main result in their analysis is that there does not exist a pure strategy symmetric Nash equilibrium in wage offers. Thus, a dispersion equilibrium must prevail. The proof of this result provided in the paper makes an error with respect to the distribution of types within a potentially deviating firm. This chapter corrects the proof and shows that the result still holds. Thus, in spite of the error, their intuition for this nonexistence result still holds. When all other firms are offering an identical wage offer, there is a discontinuity in the distribution of types in the unemployment pool, as all workers with a type below some cutoff value will be leaving firms at a lower rate than more labor-averse workers who choose to shirk. Thus, a firm can deviate by offering a higher wage, which despite the increase in costs, leads to a higher proportion of workers just above the cutoff who now want to exert effort at the higher wage. This leads to a profitable deviation and no symmetric pure strategy Nash equilibrium in wage offers can exist.

3.2 Correction of Albrecht Vroman's Proof of Proposition 3

I will employ the same notation introduced in Albrecht Vroman and denote equations repeated from Albrecht Vroman with their original numbering. The strategy of the proof provided by Albrecht and Vroman is a proof by contradiction. They assume that a common wage w^* is offered by all firm, except at most one potentially deviant firm. In this case, they derive an expression for the no-shirking cutoff within a firm offering a wage, w:

$$\theta_N(w) = \begin{cases} \frac{q[(r+a+b)w-(r+b)\omega-aw^*]}{(r+b)(r+a+b+q)e} & \text{for } w < w^* \\ \frac{q[(r+a+b+q)w-(r+b+q)\omega-aw^*]}{(r+b+q)(r+a+b+q)e} & \text{for } w \geq w^* \end{cases} \tag{A1}$$

Thus,

$$\theta_N(w^*) = q(w^* - \omega)/(r + a + b + q)e \tag{A2}$$

It is in the next step of their analysis that the authors determine "the density of θ among a potential deviant firm's employees, $g_E(\theta; w)$, is derived by equating inflows and outflows at each θ." The method of determining the outflows from a deviating firm with firm size $l(w)$ is correct. To restate their argument, "Its workers with $\theta \leq \theta_N(w)$ leave at the rate $bl(w)g_E(\theta; w)$, while its workers with $\theta > \theta_N(w)$ leave at the rate $(b + q)l(w)g_E(\theta; w)$" where $l(w)$ is the number of employees at the firm.

However, when the authors determine the inflow of workers into the deviating firm, they use the expression $a\left(\frac{l(w)}{1-u}\right)ug_U(\theta)$ where $g_U(\theta) = P(\theta|Unemployed)$. This is the rate at which an unemployed worker of type θ arrives at the deviating firm. By equating inflows and outflows, such a characterization leads to the authors writing a density

$$g_E(\theta; w) = \begin{cases} \frac{au}{b(1-u)} g_U(\theta) & \text{for } \theta \leq \theta_N(w) \\ \frac{au}{(b+q)(1-u)} g_U(\theta) & \text{for } \theta > \theta_N(w) \end{cases} \tag{A3}$$

To see why this is not in fact a density, let us follow the same (correct) steps in Albrecht Vroman to determine an expression for $g_U(\theta)$. Note that $g_U(\theta) = \left(u(\theta)g(\theta)\right)/u$, where $u(\theta)$ is the θ-specific unemployment rate and $u = \int u(\theta)dG(\theta)$ is the aggregate unemployment rate. Thus, we simply need to determine $u(\theta)$ from the steady-state conditions as in Albrecht Vroman, which yields:

$$u(\theta) = \begin{cases} \frac{b}{a+b} & \text{for } \theta \leq \theta_N(w^*) \\ \frac{b+q}{a+b+q} & \text{for } \theta > \theta_N(w^*) \end{cases} \tag{A5}$$

Thus,

$$g_U(\theta) = \begin{cases} \frac{bg(\theta)}{(a+b)u} & \text{for } \theta \leq \theta_N(w^*) \\ \frac{(b+q)g(\theta)}{(a+b+q)u} & \text{for } \theta > \theta_N(w^*) \end{cases} \tag{A6}$$

There are two cases to consider for the determination of $g_E(\theta; w)$ since the distribution will depend on whether or not the deviant firm offers a higher or lower wage than the candidate symmetric strategy of w^*. For the case in which $w \leq w^*$, we obtain the Albrecht and Vroman expression:

$$g_E(\theta; w) = \begin{cases} \frac{ag(\theta)}{(a+b)(1-u)} & \text{for } \theta \leq \theta_N(w) \\ \frac{abg(\theta)}{(b+q)(a+b)(1-u)} & \text{for } \theta_N(w) < \theta \leq \theta_N(w^*) \\ \frac{ag(\theta)}{(a+b+q)(1-u)} & \text{for } \theta > \theta_N(w^*) \end{cases} \tag{A7a}$$

This is not a well-defined probability density function in general, however, since $\int_0^\infty g_E(\theta; w)d\theta < 1$. Intuitively, they are assuming that when a firm deviates to a lower wage, for instance, the proportion of their workforce consisting of non-shirkers is unchanged, while the proportion of new shirkers falls since they now are fired more frequently. Thus, the distribution is not a density. We can see this result analytically by direct computation. First, it is helpful to determine the unemployment rate, u, from (A5).

$$u = \frac{b}{a+b}G(\theta_N(w^*)) + \frac{b+q}{a+b+q}(1 - G(\theta_N(w^*))) \tag{3.1}$$

From (A7a), we have that

$$\begin{aligned} &\int_0^\infty g_E(\theta; w)d\theta \\ &= \frac{a}{1-u}\left[\frac{1}{a+b}\left[G(\theta_N(w)) + \frac{b}{b+q}\left[G(\theta_N(w^*)) - G(\theta_N(w))\right]\right]\right. \\ &\qquad \left. + \frac{1}{a+b+q}(1 - G(\theta_N(w^*)))\right] \\ &= \frac{\frac{1}{a+b}\left[G(\theta_N(w)) + \frac{b}{b+q}\left[G(\theta_N(w^*)) - G(\theta_N(w))\right]\right] + \frac{(1-G(\theta_N(w^*)))}{a+b+q}}{\frac{G(\theta_N(w^*))}{a+b} + \frac{1-G(\theta_N(w^*))}{a+b+q}} \end{aligned}$$

where the second inequality follows from plugging in the expression in equation (3.1). Note that for an increasing cumulative distribution function, when $w < w^*$, it must be that $G(\theta_N(w)) + \frac{b}{b+q}\left[G(\theta_N(w^*)) - G(\theta_N(w))\right] < G(\theta_N(w^*))$. This implies that $\int_0^\infty g_E(\theta; w)d\theta < 1$ as claimed. A symmetric argument applies for the case in which the deviant firm chooses a higher wage.

Returning to the source of this problem, we must correct the inflow of workers from the perspective of an individual firm. The relevant inflow for a deviating firm wishing to maintain a firm size of $l(w)$ is simply the product of the density of type θ among the unemployed and the rate at which job openings are available that the firm needs to fill. This is true by the assumptions of random matching and no vacancies in the model. In particular, the inflow of type θ into the deviating firm offering a wage of w and hiring a workforce of size $l(w)$, is given by $g_U(\theta)l(w)\left(b\int_0^{\theta_N(w)} g_E(\theta;w)d\theta + (b+q)\int_{\theta_N(w)}^{\infty} g_E(\theta;w)d\theta\right)$.

Equating inflows and outflows, we have,

$$g_E(\theta;w) = \begin{cases} \frac{b\int_0^{\theta_N(w)} g_E(\theta;w)d\theta + (b+q)\int_{\theta_N(w)}^{\infty} g_E(\theta;w)d\theta}{b} g_U(\theta) & \text{for } \theta \leq \theta_N(w) \\ \frac{b\int_0^{\theta_N(w)} g_E(\theta;w)d\theta + (b+q)\int_{\theta_N(w)}^{\infty} g_E(\theta;w)d\theta}{b+q} g_U(\theta) & \text{for } \theta > \theta_N(w) \end{cases} \tag{A3$'$}$$

This expression is clearly more complicated than the one provided in AV since the density is defined in terms of itself. There is no longer a contradiction, however, in terms of the probability density function not being well-defined. In fact, imposing that $\int_0^{\infty} g_E(\theta;w)d\theta = 1$ allows us to simplify the above expression in (A3$'$). First, it is clear that we can rewrite it as

$$g_E(\theta;w) = \begin{cases} \frac{b+q\int_{\theta_N(w)}^{\infty} g_E(\theta;w)d\theta}{b} g_U(\theta) & \text{for } \theta \leq \theta_N(w) \\ \frac{b+q\int_{\theta_N(w)}^{\infty} g_E(\theta;w)d\theta}{b+q} g_U(\theta) & \text{for } \theta > \theta_N(w) \end{cases} \tag{A3$''$}$$

Next, note that from (A3$''$)

$$\int_{\theta_N(w)}^{\infty} g_E(\theta;w)d\theta = \frac{b+q\int_{\theta_N(w)}^{\infty} g_E(\theta;w)d\theta}{b+q}(1-G_U(\theta_N(w))) \tag{3.2}$$

And multiplying both sides of (3.2) by $b+q$ and rearranging terms, we have that

$$\int_{\theta_N(w)}^{\infty} g_E(\theta;w)d\theta = \frac{b(1-G_U(\theta_N(w)))}{b+qG_U(\theta_N(w))} \tag{3.3}$$

Finally, plugging (3.3) back into (A3″) yields the desired expression for the density of workers within a deviating firm:

$$g_E(\theta;w) = \begin{cases} \frac{b+q}{b+qG_U(\theta_N(w))} g_U(\theta), & \text{if } \theta \leq \theta_N(w) \\ \frac{b}{b+qG_U(\theta_N(w))} g_U(\theta), & \text{if } \theta > \theta_N(w) \end{cases} \tag{A3*}$$

and

$$G_E(\theta;w) = \begin{cases} \frac{b+q}{b+qG_U(\theta_N(w))} G_U(\theta), & \text{if } \theta \leq \theta_N(w) \\ \frac{bG_U(\theta)+qG_U(\theta_N(w))}{b+qG_U(\theta_N(w))}, & \text{if } \theta > \theta_N(w) \end{cases} \tag{3.4}$$

Expressions (A5)-(A6) are still valid and unchanged. Due to its relevance in the above expressions, it is also useful to note that the density of the unemployed population in (A6) implies a corresponding cdf given by:

$$G_U(\theta) = \begin{cases} \frac{bG(\theta)}{(a+b)u}, & \text{if } \theta \leq \theta_N(w^*) \\ \frac{bG(\theta_N(w^*))}{(a+b)u} + \frac{(b+q)(G(\theta)-G(\theta_N(w^*)))}{(a+b+q)u}, & \text{if } \theta > \theta_N(w^*) \end{cases} \tag{3.5}$$

The modified version of (A7) is now given by the following expressions

If $w \leq w^*$:

$$g_E(\theta;w) = \begin{cases} \frac{b(b+q)g(\theta)}{(b+qG_U(\theta_N(w)))(a+b)u}, & \text{if } \theta \leq \theta_N(w) \\ \frac{b^2 g(\theta)}{(b+qG_U(\theta_N(w)))(a+b)u}, & \text{if } \theta_N(w) < \theta \leq \theta_N(w^*) \\ \frac{b(b+q)g(\theta)}{(b+qG_U(\theta_N(w)))(a+b+q)u}, & \text{if } \theta > \theta_N(w^*) \end{cases} \tag{A7a*}$$

If $w \geq w^*$:

$$g_E(\theta; w) = \begin{cases} \frac{b(b+q)g(\theta)}{(b+qG_U(\theta_N(w)))(a+b)u}, & \text{if } \theta \leq \theta_N(w^*) \\ \frac{(b+q)^2 g(\theta)}{(b+qG_U(\theta_N(w)))(a+b+q)u}, & \text{if } \theta_N(w^*) < \theta \leq \theta_N(w) \\ \frac{b(b+q)g(\theta)}{(b+qG_U(\theta_N(w)))(a+b+q)u}, & \text{if } \theta > \theta_N(w) \end{cases} \qquad \text{(A7b*)}$$

To complete the proof, it is sufficient to show by the same argument provided in Albrecht Vroman that $p'^{+}(w^*) > p'^{-}(w^*)$. Note that $p(w) = G_E(\theta_N(w); w)$. When Albrecht Vroman take left and right derivatives of this function with respect to the wage, they ignore the direct effect of w on $G_E(\theta_N(w); w)$, considering only the effect of w on $\theta_N(w)$ and its subsequent effect on the cdf. This is reasonable with their specification of $G_E(\theta; w)$, in which the wage only affects the boundaries of the intervals in which θ can take values. Since I have argued that this is incorrect, it is the case as shown in (A7*) above, that we must also take into account that a change in the firm's wage offering directly affects the distribution of types in the firm. Therefore,

$$\begin{aligned} p'^{\pm}(w) &= g_E^{\pm}(\theta_N(w); w)\theta_N'^{\pm}(w) + \frac{\partial G_E(\theta_N(w); w)}{\partial w^{\pm}} \\ &= g_E^{\pm}(\theta_N(w); w)\theta_N'^{\pm}(w) - \frac{q(b+q)G_U(\theta_N(w))g_U^{\pm}(\theta_N(w))\theta_N'^{\pm}(w)}{(b+qG_U(\theta_N(w)))^2} \end{aligned}$$

Using the expressions for (A1), (A6), and (A7*), it follows that

$$
\begin{aligned}
p'^{-}(w^*) &= \theta_N'^{-}(w^*)\times \\
&\left(\frac{b(b+q)g(\theta_N(w^*))}{(b+qG_U(\theta_N(w^*)))(a+b)u} - \frac{qb(b+q)G_U(\theta_N(w^*))g(\theta_N(w^*))}{(b+qG_U(\theta_N(w^*)))^2(a+b)u}\right) \\
&= \frac{b(b+q)g(\theta_N(w^*))}{(b+qG_U(\theta_N(w^*)))(a+b)u}\left(1-\frac{qG_U(\theta_N(w^*))}{b+qG_U(\theta_N(w^*))}\right)\theta_N'^{-}(w^*) \\
&= \frac{b(b+q)g(\theta_N(w^*))}{(b+qG_U(\theta_N(w^*)))(a+b)u}\frac{q(r+a+b)}{(r+b)(r+a+b+q)e}\times \\
&\qquad\left(1-\frac{qG_U(\theta_N(w^*))}{b+qG_U(\theta_N(w^*))}\right)
\end{aligned}
$$

$$
\begin{aligned}
p'^{+}(w^*) &= \theta_N'^{+}(w^*)\times \\
&\left(\frac{(b+q)^2g(\theta_N(w^*))}{(b+qG_U(\theta_N(w^*)))(a+b+q)u} - \frac{q(b+q)^2G_U(\theta_N(w^*))g(\theta_N(w^*))}{(b+qG_U(\theta_N(w^*)))^2(a+b+q)u}\right) \\
&= \frac{(b+q)^2g(\theta_N(w^*))}{(b+qG_U(\theta_N(w^*)))(a+b+q)u}\left(1-\frac{qG_U(\theta_N(w^*))}{b+qG_U(\theta_N(w^*))}\right)\theta_N'^{+}(w^*) \\
&= \frac{(b+q)^2g(\theta_N(w^*))}{(b+qG_U(\theta_N(w^*)))(a+b+q)u}\frac{q}{(r+b+q)e}\left(1-\frac{qG_U(\theta_N(w^*))}{b+qG_U(\theta_N(w^*))}\right)
\end{aligned}
$$

By canceling common terms, checking $p'^{+}(w^*) > p'^{-}(w^*)$ is equivalent to

$$\frac{b(r+a+b)}{(a+b)(r+b)(r+a+b+q)} < \frac{(b+q)}{(a+b+q)(r+b+q)}$$

which is indeed the case. This completes the correction to the proof of Proposition 3 from the analysis in Albrecht Vroman.

3.3 Conclusion

It is clear from the steps of the proof that the result in Albrecht Vroman relies heavily on there existing a continuum of types. One can construct an analogous argument for the case with discreet types, but for the argument to go through, there must be sufficiently many similar types in the economy. With a more general specification of discrete types, we can no longer rule out a symmetric pure strategy Nash equilibrium in wage offers. In fact, an analysis in Lehr (2010) [23] shows in the simplest case with only two types of workers in the Albrecht Vroman setting that one can obtain a pure strategy symmetric equilibrium, both with or without shirking, depending on the extent of heterogeneity in the population of workers. The question of characterizing labor market equilibria with both heterogeneous workers and moral hazard constraints remains an interesting question, which Albrecht Vroman have successfully characterized in one case. It remains open to understand the role of optimal policy in such a general equilibrium setting.

Appendix A

Chapter 1

A.1 Supplemental Proofs

Proof of Proposition 5 Employing the notation introduced prior to the statement of Proposition 6, private insurers solve the same problem as in Proposition 4 given a social insurance contract, (b, t). In particular, optimal private insurance is characterized by

$$\frac{u'(c_0^I(\theta)) - u'(c_1^I(\theta))}{u'(c_1^I(\theta))} = \frac{\varepsilon_{1-a^I(\theta), b_p(\theta)}|_{b,t}}{a^I(\theta)} \tag{1.15}$$

The social planner chooses b to optimize social welfare given the budget constraint that $t = \frac{\alpha(1-\overline{a^I})+(1-\alpha)(1-\overline{a^N})}{\alpha\overline{a^I}+(1-\alpha)\overline{a^N}}$.

$$\begin{aligned} W = \alpha \int & \left(a^I(\theta) u \left(M(\theta) - t - \frac{1 - a^I(\theta)}{a^I(\theta)} b_p(\theta) \right) \right. \\ & \left. + (1 - a^I(\theta)) u(m(\theta) + b + b_p(\theta)) - h(a^I(\theta)) \right) f(\theta) d\theta \\ + (1 - \alpha) \int & \left(a^N(\theta) u \left(M(\theta) - t \right) \right. \\ & \left. + (1 - a^N(\theta)) u(m(\theta) + b) - h(a^N(\theta)) \right) f(\theta) d\theta \end{aligned}$$

Employing the envelope theorem for the agents' utility maximizing choice of a, we have that

$$\frac{dW}{db} = \alpha \int \Big((1-a^I(\theta))u'(c_0^I(\theta))(1-r(\theta)) \\ -a^I(\theta)u'(c_1^I(\theta))\frac{d}{db}\left(t+\frac{1-a^I(\theta)}{a^I(\theta)}b_p(\theta)\right)\Big) f(\theta)d\theta \\ +(1-\alpha)\int \left((1-a^N(\theta))u'(c_0^N(\theta)) - a^N(\theta)u'(c_1^N(\theta))\frac{dt}{db}\right) f(\theta)d\theta$$

Moreover,

$$\frac{d}{db}\left(t+\frac{1-a^I(\theta)}{a^I(\theta)}b_p(\theta)\right) = \frac{\alpha(1-\overline{a^I})+(1-\alpha)(1-\overline{a^N})}{\alpha\overline{a^I}+(1-\alpha)\overline{a^N}} \\ +\frac{\alpha(1-\overline{a^I})\varepsilon_{1-\overline{a^I},b}+(1-\alpha)(1-\overline{a^N})\varepsilon_{1-\overline{a^N},b}}{(\alpha\overline{a^I}+(1-\alpha)\overline{a^N})^2} \\ -r(\theta)\frac{1-a^I(\theta)}{a^I(\theta)}+\frac{b_p(\theta)}{b}\frac{1-a^I(\theta)}{a^I(\theta)^2}\varepsilon_{1-a^I(\theta),b}$$

Combining the previous two expressions and using (1.15) for optimal private insurance as in the proof of Proposition 4, it follows that

$$\frac{dW}{db} = \alpha E((1-a^I)u'(c_1^I)) + (1-\alpha)E((1-a^N)u'(c_0^N)) \\ -\frac{\alpha E(a^I u'(c_1^I))+(1-\alpha)E(a^N u'(c_1^N))}{\alpha\overline{a^I}+(1-\alpha)\overline{a^N}} \\ \times\Big[\alpha(1-\overline{a^I})+(1-\alpha)(1-\overline{a^N}) \\ +\frac{\alpha(1-\overline{a^I})\varepsilon_{1-\overline{a^I},b}+(1-\alpha)(1-\overline{a^N})\varepsilon_{1-\overline{a^N},b}}{\alpha\overline{a^I}+(1-\alpha)\overline{a^N}}\Big]$$

as desired. ■

Impact of Marginal Tax Rate on Welfare in Ex-Ante Heterogeneous Economy without Private Insurance

The marginal effect of an increase of a marginal tax rate s on welfare is given by:

$$\frac{dW}{ds} = \overline{z}\left(1 - \varepsilon_{\overline{z},1-s}\frac{s}{1-s}\right)(E(au'(c_1) + (1-a)u'(c_0)))$$
$$- E(au'(c_1)M + (1-a)u'(c_0)m) \qquad \text{(A.1)}$$

where $\overline{z} \equiv \int (a(\theta)M(\theta) + (1-a(\theta))m(\theta)) f(\theta)d\theta$ is total output and the behavioral effect of taxation on total production is given by $\varepsilon_{\overline{z},1-s}$. The optimal tax rate with no private insurance is therefore characterized by

$$\frac{s}{1-s} = \frac{1}{\varepsilon_{\overline{z},1-s}}\left[1 - \frac{1}{\overline{z}}\frac{E(au'(c_1)M + (1-a)u'(c_0)m)}{E(au'(c_1) + (1-a)u'(c_0))}\right] \qquad \text{(A.2)}$$

Proof The social planner chooses s to maximize

$$W = \int [a(\theta)u(M(\theta)(1-s) + s\overline{z})$$
$$+(1-a(\theta))u(m(\theta)(1-s) + s\overline{z}) - h(a(\theta))] f(\theta)d\theta$$

Using the envelope theorem for consumer utility maximization we have that

$$\begin{aligned}\frac{dW}{ds} = \int \Big[& a(\theta)u'(c_1(\theta))\left(\overline{z} - M(\theta) - s\frac{d\overline{z}}{d(1-s)}\right) \\ & +(1-a(\theta))u'(c_0(\theta))\left(\overline{z} - m(\theta) - s\frac{d\overline{z}}{d(1-s)}\right)\Big] f(\theta)d\theta \\ = & E(au'(c_1) + (1-a)u'(c_0))\left(\overline{z} - s\frac{d\overline{z}}{d(1-s)}\right) \\ & - E(au'(c_1)M + (1-a)u'(c_0)m) \\ = & \overline{z}\left(1 - \varepsilon_{\overline{z},1-s}\frac{s}{1-s}\right)(E(au'(c_1) + (1-a)u'(c_0))) \\ & - E(au'(c_1)M + (1-a)u'(c_0)m)\end{aligned}$$

where the last equality follows from the observation that $\frac{d\overline{z}}{d(1-s)}s = \varepsilon_{\overline{z},1-s}\frac{s}{1-s}\overline{z}$. This is the expression in (A.1) as desired and (A.2) follows immediately from setting $\frac{dW}{ds} = 0$.

■

Impact of Marginal Tax Rate on Welfare in Ex-Ante Heterogeneous Economy with Optimal Private Insurance

The marginal effect of an increase of a marginal tax rate b on welfare, taking into account the provision of optimal private insurance, is given by:

$$\begin{aligned}\frac{dW}{db} &= \overline{w}\left(1-\varepsilon_{\overline{w},1-b}\frac{b}{1-b}\right)E\left(u'(c_0)-u'(c_1)\varepsilon_{1-a,b}\frac{b_p}{b(1-r)}\right)\\ &\quad - E((1-a)u'(c_0)(m+b_p+(1-b)r))\\ &\quad - E\left(au'(c_1)\left(M-\frac{1-a}{a}\left(b_p+(1-b)r-\frac{1-b}{b}\frac{b_p}{a}\varepsilon_{1-a,b}\right)\right)\right)\end{aligned}$$

where $\overline{w} = \int\left(a(\theta)(M(\theta)-\frac{1-a(\theta)}{a(\theta)}b_p(\theta))+(1-a(\theta))(m(\theta)+b_p(\theta))\right)f(\theta)d\theta$.

Proof *Optimal Private Contract.* Firms choose $b_p(\theta)$ to maximize welfare, taking a marginal tax rate of b and rebate of $G = b\overline{w}$ as given,

$$\begin{aligned}W(\theta) &= a(\theta)u\left(\left(M(\theta)-\frac{1-a(\theta)}{a(\theta)}b_p(\theta)\right)(1-b)+G\right)\\ &\quad + (1-a(\theta))u((m(\theta)+b_p(\theta))(1-b)+G) - h(a(\theta))\end{aligned}$$

Using the envelope theorem, we have that

$$\frac{dW}{db_p(\theta)}|_{b,G} = (1-b)(1-a(\theta))u'(c_1(\theta))\left(\frac{u'(c_0(\theta))-u'(c_1(\theta))}{u'(c_1(\theta))} - \frac{\varepsilon_{1-a(\theta),b_p(\theta)}|_{b,G}}{a(\theta)}\right)$$

This expression is the analog of (1.14) with the only change being the multiplicative term of $(1-b)$. The optimal private insurance contract, conditional on $b \neq 1$, is therefore characterized by

$$\frac{u'(c_0(\theta))-u'(c_1(\theta))}{u'(c_1(\theta))} = \frac{\varepsilon_{1-a(\theta),b_p(\theta)}|_{b,G}}{a(\theta)} \tag{A.3}$$

Note that this expression is identical to that in (1.15) for the optimal private contract given tax policy.

Social Insurance in the Presence of Optimized Private Insurance. The government chooses b to maximize social welfare, taking into account $b_p(\theta)$ set as in (A.3). In particular, welfare is given by

$$W = \int \Big[a(\theta) u \left(\left(M(\theta) - \frac{1-a(\theta)}{a(\theta)} b_p(\theta) \right) (1-b) + b\overline{w} \right)$$
$$+ (1-a(\theta)) u((m(\theta) + b_p(\theta))(1-b) + b\overline{w}) - h(a(\theta)) \Big] f(\theta) d\theta$$

Employing the envelope theorem, we have that

$$\frac{dW}{db} = \int \Big[a(\theta) u'(c_1(\theta)) \left(\overline{w} - b\frac{d\overline{w}}{d(1-b)} - M(\theta) + \frac{1-a(\theta)}{a(\theta)} b_p(\theta) \right.$$
$$\left. -(1-b)\left(-\frac{1-a(\theta)}{a(\theta)} r(\theta) + b_p(\theta) \frac{d\frac{1-a(\theta)}{a(\theta)}}{db} \right) \right)$$
$$+ (1-a(\theta)) u'(c_0(\theta)) \left(\overline{w} - b\frac{d\overline{w}}{d(1-b)} - m(\theta) - b_p(\theta) \right.$$
$$\left. -(1-b) r(\theta) \right) \Big] f(\theta) d\theta$$

Note that $\frac{d\overline{w}}{d(1-b)} b = \varepsilon_{\overline{w},1-b} \frac{b}{1-b} \overline{w}$ and $\frac{d\frac{1-a(\theta)}{a(\theta)}}{db} = \varepsilon_{1-a(\theta),b} \frac{1-a(\theta)}{ba(\theta)^2}$. This implies

$$\frac{dW}{db} = \overline{w} \left(1 - \varepsilon_{\overline{w},1-b} \frac{b}{1-b} \right) E(au'(c_1) + (1-a)u'(c_0))$$
$$- E((1-a)u'(c_0)(m + b_p + (1-b)r))$$
$$- E\left(au'(c_1) \left(M - \frac{1-a}{a} \left(b_p + (1-b)r - \frac{1-b}{b}\frac{b_p}{a} \varepsilon_{1-a,b} \right) \right) \right)$$

From (A.3) we know that

$$E(a(u'(c_1) - u'(c_0))) = -E(u'(c_1)\varepsilon_{1-a,b_p}|_{b,G}) = -E\left(u'(c_1)\varepsilon_{1-a,b} \frac{b_p}{b(1-r)} \right)$$

Thus,

$$\begin{aligned}\frac{dW}{db} &= \overline{w}\left(1-\varepsilon_{\overline{w},1-b}\frac{b}{1-b}\right)E\left(u'(c_0)-u'(c_1)\varepsilon_{1-a,b}\frac{b_p}{b(1-r)}\right)\\ &- E((1-a)u'(c_0)(m+b_p+(1-b)r))\\ &- E\left(au'(c_1)\left(M-\frac{1-a}{a}\left(b_p+(1-b)r-\frac{1-b}{b}\frac{b_p}{a}\varepsilon_{1-a,b}\right)\right)\right)\end{aligned}$$

as claimed. ■

A.2 Supplemental Figures for Simulation in Figure 1-3

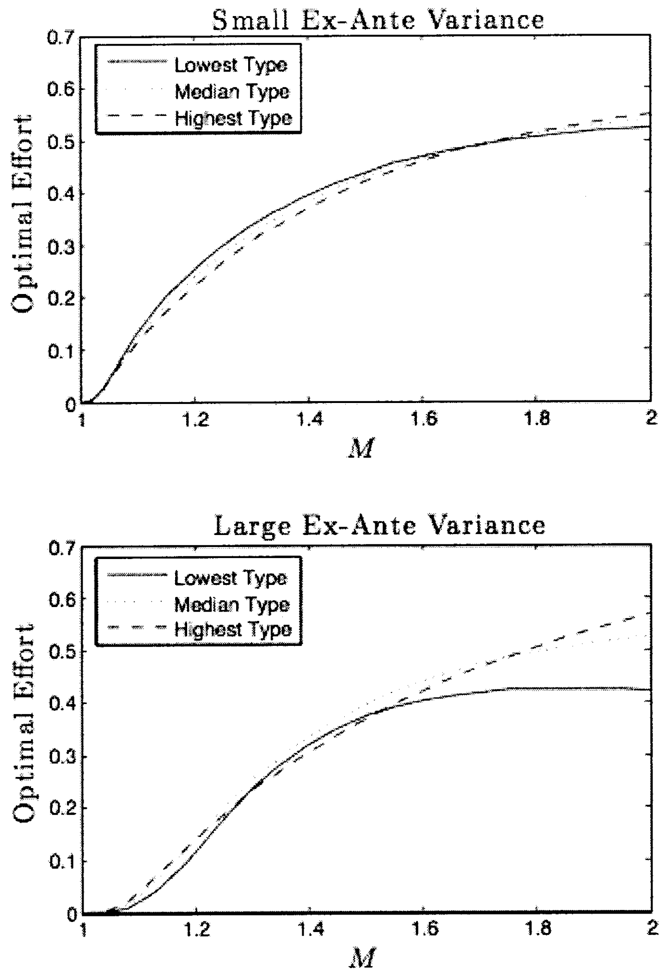

Figure A-1: Optimal Effort with No Private Insurance

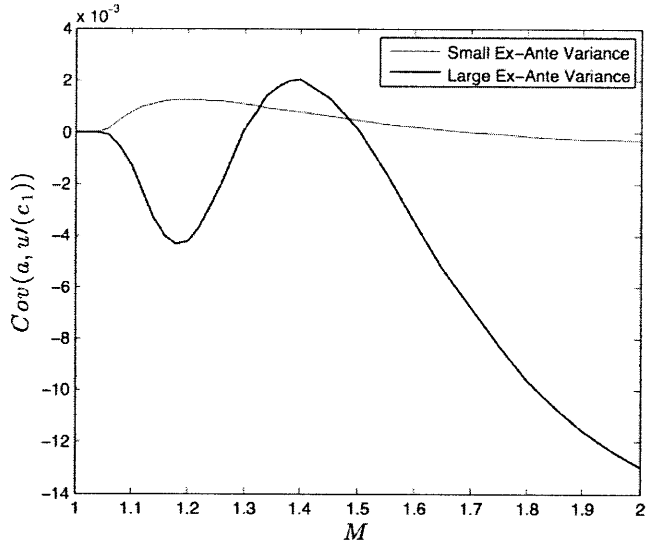

Figure A-2: Effort-Marginal Utility Covariance with No Private Insurance

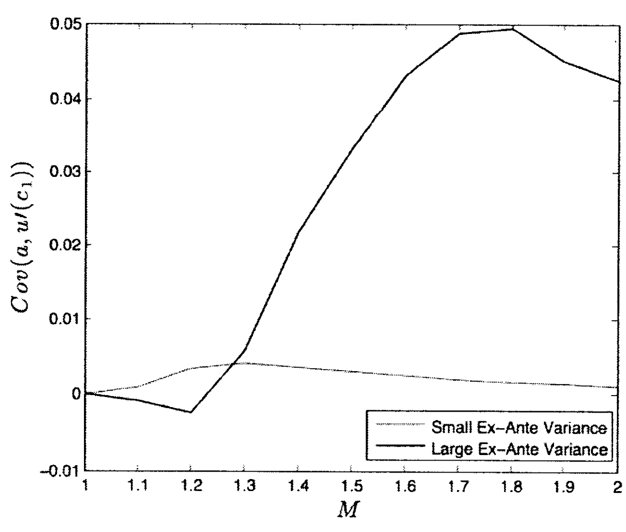

Figure A-3: Effort-Marginal Utility Covariance with Private Insurance

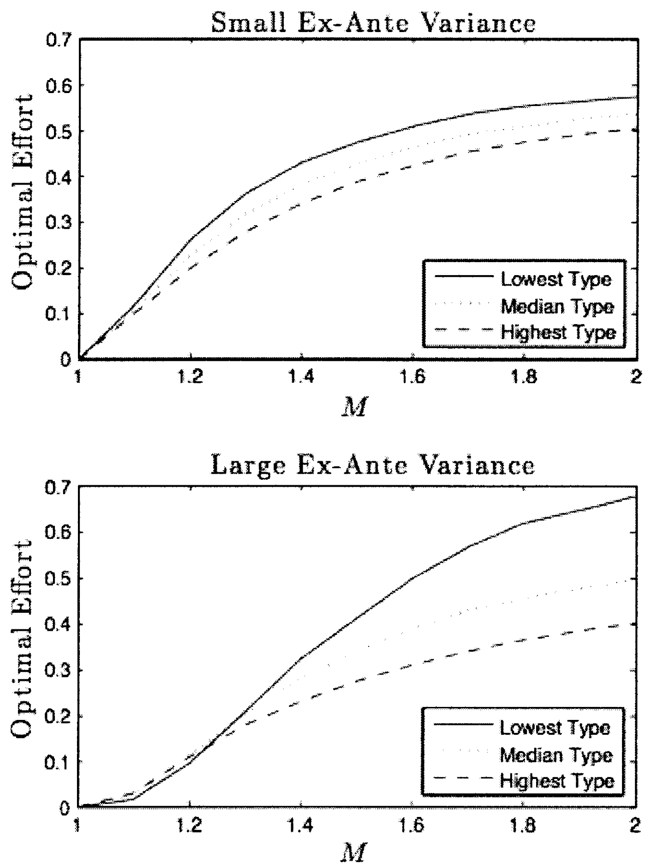

Figure A-4: Optimal Effort with Private Insurance

Appendix B

Chapter 2

Proof of Proposition 2 Part (a) follows immediately from the work preceding the statement of the Proposition and the results in (2.14) and (2.23). From (2.14) we know that $\frac{\theta_h}{\theta_l} \geq \psi$ is sufficient to guarantee the desired equilibrium. Moreover, it is the unique symmetric Nash equilibrium since the only other candidate would require (2.23) to be satisfied, which will not be the case for any accession rate, $a > 0$, when $\frac{\theta_h}{\theta_l} \geq \psi$.

To prove part (b), let us begin by noting that the wages given by (2.8) and (2.18) constitute a pure strategy symmetric Nash equilibrium, with non-shirking and shirking low type workers, respectively, if and only if firms are earning zero profits and have no profitable deviation, given all other firms offering the same wage contract. In particular, a non-shirking PSSNE obtains precisely when the value function in (2.7) equals zero and (2.13) is satisfied. Similarly, a shirking PSSNE obtains when the value function in (2.17) equals zero and (2.23) is satisfied. Note that for a given fixed cost, $c > 0$, there will be accession rates a_1 and a_2 which yield zero profits in (2.7) and (2.17), respectively.

Moreover, the accession rates as functions of c are strictly decreasing. This follows from inspection of the profit functions in (2.7) and (2.17), in conjunction with the optimal wages paid. The wages paid in both a non-shirking and shirking PSSNE are strictly increasing in the accession rate, as seen from (2.8) and (2.18). An increase in

c strictly reduces profits. To maintain zero equilibrium profits in (2.7), this requires that the wage and therefore a_1 decrease. For the shirking case, to maintain zero equilibrium profits in (2.17) we similarly need a_2 to decrease, which will lower the total labor costs and also increase the share of productive workers in the firm in steady state. Since a must be strictly positive in equilibrium, we will restrict our attention to costs, c, that are not sufficiently large so as to violate $a_i(c) > 0$ for either $i = 1, 2$.

Given a cost, c, and the equilibrium accession rates, we must also verify that there is no profitable deviation for a firm in the proposed PSSNE. Let us define the following functions:

$$\Gamma_1(c) = \frac{\theta_h}{\theta_l} - \frac{\pi(s+d)}{s+\pi d}\frac{r+a_1(c)+s}{r+s} + \frac{a_1(c)}{r+s} \tag{B.1}$$

$$\Gamma_2(c) = \frac{\theta_l}{\theta_h} - \frac{a_2(c)+s+d\pi}{\pi(a_2(c)+s+d)} - \frac{a_2(c)(a_2(c)+s)(1-\pi)}{\pi(a_2(c)+s+d)(r+s+d)} \tag{B.2}$$

which are simply a rearrangement of the no profitable deviation conditions in (2.13) and (2.23) so that there is no profitable deviation if and only iff $\Gamma_i(c) \geq 0$ where $i = 1$ corresponds to the no-shirking PSSNE and $i = 2$ to the shirking PSSNE.

To complete the proof, we will establish a series of facts about $\Gamma_i(c)$ for $i = 1, 2$.

Fact 1 Since $a_i(c)$ are strictly decreasing in c, it follows that $\Gamma_1(c)$ is strictly decreasing in c, while $\Gamma_2(c)$ is strictly increasing in c.

Fact 2 Claim that $\Gamma_1(c) > 0$ for $c > 0$ sufficiently small. This follows from (B.1) in which $\lim_{c\to 0} a_1(c) = \infty$. Then, since the right hand side of (B.1) is linearly increasing in $a_1(c)$, we have that $\lim_{c\to 0}\Gamma_1(c) = \infty$, establishing our claim.

Fact 3 Claim that there exists a unique cost, c^* such that $\Gamma_2(c^*) = 0$. In particular, c^* is the cost such that the corresponding accession rate, $a_2^* \equiv a_2(c^*)$ makes firms in a potential shirking PSSNE weakly prefer the candidate equilibrium wage, (2.18), to deviating to the best wage offer which induces effort from all workers. To see this, note that $\Gamma_2(c^*) = 0$ is equivalent to

$$a_2^{*2}\theta_h(1-\pi) + a_2^*\left(\theta_h s(1-\pi) + (r+s+d)(\theta_h - \pi\theta_l)\right) \\ + (r+s+d)\left(\theta_h(s+d\pi) - \pi(s+d)\theta_l\right) = 0 \tag{B.3}$$

Thus, a_2^* is a root of the quadratic

$$Q(x) = x^2\theta_h(1-\pi) + x\left(\theta_h s(1-\pi) + (r+s+d)(\theta_h - \pi\theta_l)\right) \\ + (r+s+d)\left(\theta_h(s+d\pi) - \pi(s+d)\theta_l\right) \tag{B.4}$$

It is clear that $Q(x)$ is strictly convex and $Q(0) < 0$ since it is assumed that $\frac{\theta_h}{\theta_l} < \frac{\pi(s+d)}{s+\pi d}$. Thus, there exists a unique strictly positive root, which we denote by a_2^*, and $c^* = \max_n f\left(\frac{\pi(a_2^*+s+d)}{a_2^*+s+\pi d}n\right) - \frac{r+a_2^*+s+d}{d}\theta_h n$, establishing the claim.

Fact 4 Claim that $\Gamma_1(c^*) < 0$. Let us first express $a_1(c^*)$ in terms of a_2^*. $\Gamma_2(c^*) = 0$ is equivalent to

$$\frac{a_2^*\theta_h + (r+s+d)\theta_l}{d} = \frac{(r+a_2^*+s+d)(a_2^*+s+\pi d)}{d\pi(a_2^*+s+d)}\theta_h \tag{B.5}$$

In addition, since profits are equal in the two possible PSSNE, namely they are both zero, we have that

$$\max_n f(n) - \frac{r+a_1(c^*)+s+d}{d}\theta_l n = \max_n f\left(\frac{\pi(a_2^*+s+d)}{a_2^*+s+\pi d}n\right) - \frac{r+a_2^*+s+d}{d}\theta_h n$$

which is equivalent to the effective cost of labor being equal across these two cases, i.e.,

$$\frac{r+a_1(c^*)+s+d}{d}\theta_l = \frac{(r+a_2^*+s+d)(a_2^*+s+\pi d)}{d\pi(a_2^*+s+d)}\theta_h \tag{B.6}$$

Combining (B.5) and (B.6) we have that

$$\frac{r + a_1(c^*) + s + d}{d}\theta_l = \frac{a_2^*\theta_h + (r + s + d)\theta_l}{d}$$

which implies

$$a_1(c^*) = \frac{d}{\theta_l}\frac{a_2^*\theta_h + (r + s + d)\theta_l}{d} - r - s - d = a_2^*\frac{\theta_h}{\theta_l} \tag{B.7}$$

Next, note that $\Gamma_1(c^*) < 0$ is equivalent to

$$\begin{aligned}\frac{\pi(s+d)}{s+\pi d}\theta_l > &\ \frac{(r+s)\theta_h + a_1(c^*)\theta_l}{r+s+a_1(c^*)} \\ = &\ \frac{(r+s+a_2^*)\theta_h\theta_l}{a_2^*\theta_h + (r+s)\theta_l} \text{ by (B.7)}\end{aligned} \tag{B.8}$$

which can be equivalently expressed with some algebraic manipulation as

$$a_2^* < \frac{(r+s)\pi(s+d)}{\theta_h s(1-\pi)}\left(\theta_l - \frac{(s+\pi d)\theta_h}{\pi(s+d)}\right) \equiv \kappa \tag{B.9}$$

where $\kappa > 0$ since by assumption $\frac{\theta_h}{\theta_l} < \frac{\pi(s+d)}{s+\pi d}$. Moreover, since $Q(x)$ is a strictly convex quadratic with $Q(0) < 0$ and $Q(a_2^*) = 0$, to establish that $\Gamma_1(c^*) < 0$ it is sufficient to show that $Q(\kappa) > 0$. The algebra to establish this fact is somewhat tedious and details are available from the author, but intermediate steps are:

$$\begin{aligned}&Q(\kappa) > 0 \\ \Leftrightarrow &\ \frac{(r+s)^2\pi(s+d)}{s^2\theta_h(1-pi)}\left(\theta_l - \frac{(s+\pi d)\theta_h}{\pi(s+d)}\right) + \frac{(r+s+d)(r+s)(\theta_h - \pi\theta_l)}{\theta_h s(1-\pi)} - d > 0 \\ \Leftrightarrow &\ (r+s)^2(\pi(s+d)\theta_l - (s+\pi d)\theta_h) + s(\theta_h - \pi\theta_l)(dr + (r+s)^2) \\ &\qquad + ds^2\pi(\theta_h - \theta_l) > 0 \\ \Leftrightarrow &\ ((r+s)^2 - s^2)\pi(\theta_l - \theta_h) + sr(\theta_h - \pi\theta_l) > 0 \\ \Leftrightarrow &\ \pi(r+s)(\theta_l - \theta_h) + s\theta_h(1-\pi) > 0\end{aligned}$$

which is true. This establishes the desired claim.

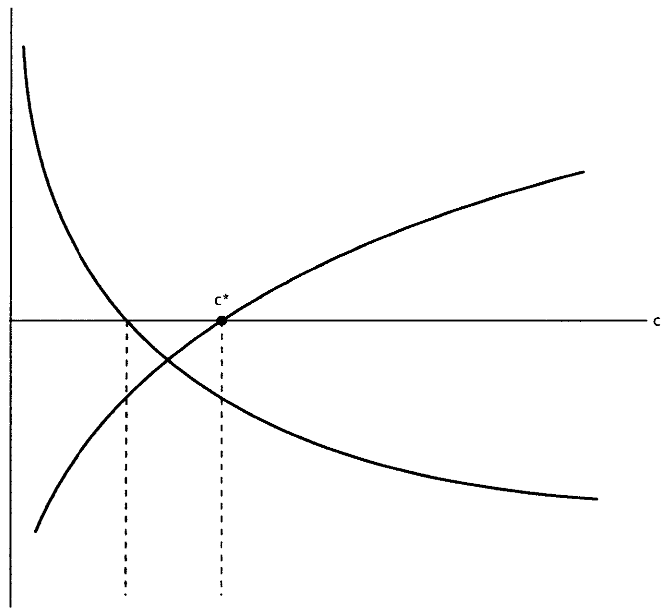

Figure B-1: Existence of PSSNE

We can now assemble Facts 1- 4 to see that the statement of part (b) of Proposition 2 follows directly. It is easiest to understand the collection of facts in a figure.

Figure B-1 plots $\Gamma_1(c)$ and $\Gamma_2(c)$. We know that a pure strategy symmetric Nash equilibrium exists iff $\Gamma_i(c) \geq 0$ where $i = 1$ corresponds to no shirking by all workers and $i = 2$ corresponds to shirking by the low type workers in equilibrium. The monotonicity of the functions follows from Fact 1. Fact 3 establishes that $\Gamma_2(c)$ crosses the $c-$axis. Facts 2 and 4 establish that there exist values of c such that $\Gamma_1(c)$ takes on both positive and negative values and in particular that $\Gamma_1(c)$ is negative when $\Gamma_2(c) = 0$. Taken together, we have a picture which shows that for small values of c, a unique PSSNE exists without shirking, for intermediate c, no PSSNE exists, and for sufficiently large c, a unique PSSNE exists with shirking. This completes the proof. ■

Proof of Proposition 3 We prove this result by showing that when (2.13) is satisfied, the best possible deviation which solves (2.25) given that all other firms are offering the wage contract (2.8) is for firms to offer (2.8). Thus, the availability of offering a two-part contract is not optimal and thus does not affect the proposed no shirking PSSNE. We will proceed in steps. First we will simplify the constraint set. Then we will characterize the solution to (2.25) with a reduced constraint set. Finally we will verify that the dropped constraints are satisfied at the optimum.

Step 1 Guess that the relevant constraints for a firm facing problem (2.25) are (2.26′), (2.30′), and a monotonicity constraint, $L_h \geq L_l$. Constraints (2.26′) and (2.30′) will bind at the optimum as the firm wants the lowest feasible income payments. Constraint (2.26′) binding with the monotonicity constraint imply that (2.29′) is satisfied. Solving out the binding constraints, we obtain

$$y_h = \frac{r+s+d}{d}(\theta_l L_l + \theta_h(L_h - L_l)) + \frac{a}{d}\theta_l \tag{B.10}$$

$$y_l = \frac{r+s+d}{d}\theta_l L_l + \frac{r+s}{d}\theta_h(L_h - L_l) + \frac{a}{d}\theta_l \tag{B.11}$$

The monotonicity constraint with (B.11) imply that (2.31′) is also satisfied. We will continue to solve the problem and will check that (2.27′) and (2.28′) are satisfied at the optimum.

Step 2 The problem for the firm with the new constraint set is to solve

$$\begin{aligned}
\max_{L_l, L_h, n} \; & f(n(\pi L_h + (1-\pi)L_l)) \\
& - n\left[L_h\theta_h\left(\frac{r+s}{d} + \pi\right) + L_l\left(\frac{r+s}{d}(\theta_l - \theta_h) + \theta_l - \pi\theta_h\right)\right] \\
& - n\frac{a}{d}\theta_l - c
\end{aligned}$$

$$s.t. \quad L_h \leq 1 \tag{B.12}$$

$$L_l \leq L_h \tag{B.13}$$

$$h \leq L_l \tag{B.14}$$

Letting the Lagrange multipliers on the constraints be λ_1, λ_2, and λ_3, respectively,

the first order conditions (which are necessary and sufficient) for this problem are

$$[L_h]: f'(\cdot)n\pi = n\theta_h\left(\frac{r+s}{d}+\pi\right)+\lambda_1-\lambda_2 \tag{B.15}$$

$$[L_l]: f'(\cdot)n(1-\pi) = n\left(\frac{r+s}{d}(\theta_l-\theta_h)+\theta_l-\pi\theta_h\right)+\lambda_2-\lambda_3 \tag{B.16}$$

$$\begin{aligned}[n]: f'(\cdot)(\pi L_h+(1-\pi)L_l) &= L_h\theta_h\left(\frac{r+s}{d}+\pi\right)\\ &+ L_l\left(\frac{r+s}{d}(\theta_l-\theta_h)+\theta_l-\pi\theta_h\right)+\frac{a}{d}\theta_l\end{aligned} \tag{B.17}$$

The first order conditions imply that

$$L_h(\lambda_1-\lambda_2)+L_l(\lambda_2-\lambda_3)=n\frac{a}{d}\theta_l \tag{B.18}$$

Note that $\lambda_1>0$. Otherwise, from (B.18) we have $\lambda_2(L_l-L_h)-\lambda_3 L_l>0$, which is a contradiction since $\lambda_i\geq 0$ for all i. Thus, with $\lambda_1>0$, we have that $L_h=1$. Hence, $\lambda_1-\lambda_2=n\frac{a}{d}\theta_l-L_l(\lambda_2-\lambda_3)$. Plugging this expression for $\lambda_1-\lambda_2$ into (B.15) and dividing (B.15) by (B.16), we have

$$\frac{\pi}{1-\pi}=\frac{\theta_h\left(\frac{r+s}{d}+\pi\right)+\frac{a}{d}\theta_l-\frac{L_l}{n}(\lambda_2-\lambda_3)}{\frac{r+s}{d}(\theta_l-\theta_h)+\theta_l-\pi\theta_h+\frac{\lambda_2-\lambda_3}{n}} \tag{B.19}$$

Moreover, with some algebra, one can establish the fact that

$$\frac{\pi(s+d)}{s+\pi d}\frac{r+a+s}{r+s}-\frac{a}{r+s}>\frac{\pi(r+a+s+d)-a}{r+s+d\pi} \tag{B.20}$$

And by (2.13), which guarantees no profitable deviation to a shirking contract, we have that

$$\frac{\theta_h}{\theta_l}\geq\frac{\pi(s+d)}{s+\pi d}-\frac{as(1-\pi)}{(s+\pi d)(r+s)} \tag{2.13}$$

Combining (B.20) and (2.13), we have that

$$\frac{\theta_h}{\theta_l} > \frac{\pi(r+a+s+d)-a}{r+s+d\pi}$$

$$\Leftrightarrow$$

$$\frac{\pi}{1-\pi} < \frac{\theta_h\left(\frac{r+s}{d}+\pi\right)+\frac{a}{d}\theta_l}{\frac{r+s}{d}(\theta_l-\theta_h)+\theta_l-\pi\theta_h} \tag{B.21}$$

Constraint (B.21) implies that in order for (B.19) to hold, we must have $\lambda_2 - \lambda_3 > 0$, in which case $\lambda_2 > 0$ and $\lambda_3 = 0$, implying that $L_l = 1$. Thus, the best response for a firm is to offer a single labor contract with wage exactly equal to the equilibrium wage in (2.8).

Step 3 It remains to verify that dropping (2.27′) and (2.28′) was legitimate. Since the solution is one in which the firm offers a single contract to all workers, (2.27′) and (2.28′) are equivalent. Thus checking one of them is equivalent to:

$$\frac{r+a+s+d}{d}\theta_l \geq \frac{r+s+d}{d}\theta_h + \frac{a}{r+a+s}\left(\frac{r+a+s+d}{d}\theta_l - \theta_h\right)$$

$$\Leftrightarrow$$

$$\frac{r+s+d}{d}(\theta_l-\theta_h) \geq \frac{a}{r+a+s}(\theta_l-\theta_h)$$

which is true. Moreover, since the best response is the equilibrium wage, we know from Proposition 1 that no worker will reject the wage offer, validating the setup of the problem in which the deviating firm ignored the potential for workers to reject the contract offered in the best deviation. This completes the proof. ■

Proof of Lemma 1 We prove this result by construction and observe that (2.34) is the relevant condition for determining the optimal choice of L_l. We will proceed in steps. First we will simplify the constraint set. Then we will characterize the solution to (2.25) with a reduced constraint set. Finally we will verify that the dropped constraints are satisfied at the optimum.

Step 1 Guess that the relevant constraints for a firm facing problem (2.33) are (2.26″), (2.30″), and a monotonicity constraint, $L_h \geq L_l$. Constraints (2.26″) and (2.30″) will bind at the optimum as the firm wants the lowest feasible income payments. Constraint (2.26″) binding with the monotonicity constraint imply that (2.29″) is satisfied. Solving out the binding constraints, we obtain

$$y_h = \frac{r+s+d}{d}(\theta_h L_h + (\theta_l - \theta_h)L_l) + \frac{a}{d}\theta_h \tag{B.22}$$

$$y_l = \frac{r+s+d}{d}\theta_l L_l + \frac{r+s}{d}\theta_h(L_h - L_l) + \frac{a}{d}\theta_h \tag{B.23}$$

The monotonicity constraint with (B.23) imply that (2.31″) is also satisfied. We will continue to solve the problem and will check that (2.27″) and (2.28″) are satisfied at the optimum.

Step 2 The problem for the firm with the new constraint set is to solve

$$\begin{aligned}
&\max_{L_l, L_h, n} f(n(\widetilde{\psi} L_h + (1-\widetilde{\psi})L_l)) \\
&\qquad - n\left[L_h\theta_h\left(\frac{r+s}{d} + \widetilde{\psi}\right) + L_l\left(\frac{r+s}{d}(\theta_l - \theta_h) + \theta_l - \widetilde{\psi}\theta_h\right)\right] \\
&\qquad - n\frac{a}{d}\theta_h - c
\end{aligned}$$

$$s.t. \quad L_h \leq 1 \tag{B.24}$$

$$L_l \leq L_h \tag{B.25}$$

$$h \leq L_l \tag{B.26}$$

Letting the Lagrange multipliers on the constraints be λ_1, λ_2, and λ_3, respectively,

the first order conditions (which are necessary and sufficient) for this problem are

$$[L_h] : f'(\cdot)n\widetilde{\psi} = n\theta_h\left(\frac{r+s}{d} + \widetilde{\psi}\right) + \lambda_1 - \lambda_2 \tag{B.27}$$

$$[L_l] : f'(\cdot)n(1-\widetilde{\psi}) = n\left(\frac{r+s}{d}(\theta_l - \theta_h) + \theta_l - \widetilde{\psi}\theta_h\right) + \lambda_2 - \lambda_3 \tag{B.28}$$

$$\begin{aligned}[n] : f'(\cdot)(\widetilde{\psi}L_h + (1-\widetilde{\psi})L_l) = {} & L_h\theta_h\left(\frac{r+s}{d} + \widetilde{\psi}\right) \\ & + L_l\left(\frac{r+s}{d}(\theta_l - \theta_h) + \theta_l - \widetilde{\psi}\theta_h\right) + \frac{a}{d}\theta_h\end{aligned} \tag{B.29}$$

The first order conditions imply that

$$L_h(\lambda_1 - \lambda_2) + L_l(\lambda_2 - \lambda_3) = n\frac{a}{d}\theta_h \tag{B.30}$$

Note that $\lambda_1 > 0$. Otherwise, from (B.30) we have $\lambda_2(L_l - L_h) - \lambda_3 L_l > 0$, which is a contradiction since $\lambda_i \geq 0$ for all i. Thus, with $\lambda_1 > 0$, we have that $L_h = 1$. Hence, $\lambda_1 - \lambda_2 = n\frac{a}{d}\theta_h - L_l(\lambda_2 - \lambda_3)$. Plugging this expression for $\lambda_1 - \lambda_2$ into (B.27) and dividing (B.27) by (B.28), we have

$$\frac{\widetilde{\psi}}{1-\widetilde{\psi}} = \frac{\theta_h\left(\frac{r+s}{d} + \widetilde{\psi}\right) + \frac{a}{d}\theta_h - \frac{L_l}{n}(\lambda_2 - \lambda_3)}{\frac{r+s}{d}(\theta_l - \theta_h) + \theta_l - \widetilde{\psi}\theta_h + \frac{\lambda_2 - \lambda_3}{n}} \tag{B.31}$$

Note that we cannot have both λ_2 and λ_3 strictly positive. Thus, if $\lambda_2 - \lambda_3 > 0$, then $\lambda_2 > 0$ and $\lambda_3 = 0$, implying that $L_l = 1$, whereas if $\lambda_2 - \lambda_3 < 0$, then $\lambda_2 = 0$ and $\lambda_3 > 0$, implying that $L_l = h$. For $\lambda_2 - \lambda_3 = 0$, any feasible choice of L_l is optimal. These facts immediately imply that the optimal menu of contracts depends on the relationship of the following inequality:

$$\frac{\widetilde{\psi}}{1-\widetilde{\psi}} \gtreqless \frac{\theta_h\left(\frac{r+s}{d} + \widetilde{\psi}\right) + \frac{a}{d}\theta_h}{\frac{r+s}{d}(\theta_l - \theta_h) + \theta_l - \widetilde{\psi}\theta_h} \tag{B.32}$$

as characterized in the statement of the lemma.

Step 3 Finally, it remains to verify that the constraints (2.27″) and (2.28″) are satisfied at the optimum in both cases. Starting with the two-part contracts, we see that by employing the characterization of the labor contracts in the lemma, constraint (2.27″) is equivalent to

$$\frac{a}{d}\theta_h + \frac{r+s+d}{d}L_l(\theta_l - \theta_h) \geq \frac{a}{d}\theta_h$$

which is true. To verify constraint (2.28″), note that it is equivalent to

$$\begin{aligned}
&(r+s+d)\left(\frac{r+a+s}{d}\theta_h + \frac{r+s+d}{d}(\theta_l - \theta_h)L_l\right) \\
&\qquad \geq (r+s)\left(\frac{r+a+s}{d}\theta_h + \frac{r+s}{d}(\theta_l - \theta_h)L_l + \theta_l L_l\right) + a\theta_h \\
&\qquad \Longleftrightarrow (r+s)\theta_h + L_l(\theta_l - \theta_h)\left(\frac{(r+s+d)^2}{d} - \frac{(r+s)^2)}{d}\right) \geq (r+s)\theta_l L_l \\
&\qquad \Longleftrightarrow (r+s)\theta_h + L_l(\theta_l(r+s+d) - \theta_h(d + 2(r+s))) \geq 0 \\
&\qquad \Leftarrow \quad (r+s)\theta_h L_l + L_l(\theta_l(r+s+d) - \theta_h(d+2(r+s))) \geq 0 \\
&\qquad \Longleftrightarrow L_l(\theta_l - \theta_h)(r+s+d) \geq 0
\end{aligned}$$

which is true. Finally, we must verify that in the case of a pooling contract that (2.27″) and (2.28″) are satisfied at the optimum. These constraints are equivalent with a pooling contract, so it is sufficient to verify that

$$\frac{a\theta_h + (r+s+d)\theta_l}{d} \geq \frac{a}{d}\theta_h$$

which is clearly true. This completes the proof. ■

Proof of Lemma 2 For the high types we have that

$$\begin{aligned} y_h - \theta_h L_h &\geq \frac{r+s}{d}\theta_h L_h + \frac{a}{d}\theta_h \text{ by (2.27'')} \\ &\geq \frac{a}{d}\theta_h \\ &= rU(\theta_h) \text{ by (2.32)} \\ \Leftrightarrow V_N\left(\frac{y_h}{L_h}, L_h; \theta_h\right) &\geq U(\theta_h) \end{aligned}$$

And for the low types,

$$\begin{aligned} V_N\left(\frac{y_l}{L_l}, L_l; \theta_l\right) &\geq U(\theta_l) \\ \Leftrightarrow y_l - \theta_l L_l &\geq rU(\theta_l) \\ &= \frac{a}{d}\theta_h \text{ by (2.32)} \end{aligned}$$

By Lemma 1 we know that $y_l = \frac{a}{d}\theta_h + \frac{r+s+d}{d}\theta_l L_l + \frac{r+s}{d}\theta_h(1-L_l)$ for $L_l \in [h, 1]$. Thus, for any feasible L_l, the desired inequality is

$$\frac{r+s}{d}\left(\theta_l L_l + \theta_h(1-L_l)\right) \geq 0$$

which is true. ■

Proof of Proposition 4 From Lemma 1 we know the best possible deviations for firms considering offering contracts which induce effort from all workers. There are two cases to consider depending on whether this deviation is a contract in which it is optimal for the labor hours of the low type to be at the maximum or minimum number of feasible hours. Note that we can ignore the case in which there is a strictly interior solution for L_l since in such a case a corner solution yields the same profit level, and the profit is the only relevant statistic in the proof. We take each case in turn.

Case 1: $L_l = h$

In this case, in order to prevent a profitable deviation, we must have that

$$\begin{aligned}
&\max_n f\left(\frac{\pi(a+s+d)}{a+s+\pi d}n\right) - \frac{r+a+s+d}{d}\theta_h n - c \\
&\geq \max_n f(n(\widetilde{\psi}+(1-\widetilde{\psi})h)) - n\theta_h\left(\frac{r+s}{d}+\widetilde{\psi}\right) \\
&\quad - n\left(h\left(\frac{r+s}{d}(\theta_l-\theta_h)+\theta_l-\widetilde{\psi}\theta_h\right)+\frac{a}{d}\theta_h\right) - c
\end{aligned}$$

or equivalently,

$$R(h) \geq S \tag{B.33}$$

where

$$\begin{aligned}
R(h) &\equiv \frac{\theta_h\left(\frac{r+a+s}{d}+\widetilde{\psi}\right)+h\left(\frac{r+s}{d}(\theta_l-\theta_h)+\theta_l-\widetilde{\psi}\theta_h\right)}{\widetilde{\psi}+(1-\widetilde{\psi})h} \\
S &\equiv \frac{\frac{r+a+s+d}{d}\theta_h}{\frac{\pi(a+s+d)}{a+s+\pi d}}
\end{aligned}$$

First observe that:

$$\begin{aligned}
&R(0) > S \\
&\Leftrightarrow \frac{r+a+s+\widetilde{\psi}d}{r+a+s+d} > \widetilde{\psi}\frac{a+s+d\pi}{\pi(a+s+d)} = \frac{s(a+s+d\pi)}{s(a+s+d)+ad(1-\pi)} \\
&\Leftrightarrow (1-\pi)(sr+a(r+a+s)) > 0
\end{aligned}$$

which is true, where the last inequality follows from simple algebraic manipulation.

Next, note that

$$R'(h) = \frac{\widetilde{\psi}(\theta_l-\theta_h)\frac{r+s+d}{d} - (1-\widetilde{\psi})\frac{r+a+s}{d}\theta_h}{(\widetilde{\psi}+(1-\widetilde{\psi})h)^2} \tag{B.34}$$

and therefore

$$R'(h) \geq 0 \Leftrightarrow \frac{\widetilde{\psi}}{1-\widetilde{\psi}} \geq \frac{(r+a+s)\theta_h}{(r+s+d)(\theta_l - \theta_h)} \tag{B.35}$$

Since we are in the case of a separating contract, however, we know that it must be that (2.34) holds, so in particular,

$$\frac{\widetilde{\psi}}{1-\widetilde{\psi}} \geq \frac{\theta_h \left(\frac{r+a+s}{d} + \widetilde{\psi}\right)}{\frac{r+s}{d}(\theta_l - \theta_h) + \theta_l - \widetilde{\psi}\theta_h} = \frac{(r+a+s)\theta_h}{(r+s+d)(\theta_l - \theta_h)}$$

where the equality follows from plugging in the expression for $\widetilde{\psi}$. This is exactly the condition in (B.35). Thus, since $R(0) > S$ and $R'(h) \geq 0$, it follows that $R(h) > S$ for all h. This guarantees no profitable deviation in this case.

Case 2: $L_l = H = 1$

If the optimal deviation is to offer a single type of contract as in Lemma 1, then the no deviation condition is simply (2.23) since there are no variable hours in the proposed equilibrium or deviation. Moreover, (2.35) must hold, which imposes an upper bound on $\frac{\theta_l}{\theta_h}$ which is strictly greater (for $\pi < 1$) than the lower bound provided by (2.23), so there exists a range of $\frac{\theta_l}{\theta_h}$ in which the best deviation is a pooling contract, but such a deviation is not profitable. The range is given by

$$\frac{\theta_l}{\theta_h} \in \left[\frac{a+s+d\pi}{\pi(a+s+d)} + \frac{a(a+s)(1-\pi)}{\pi(a+s+d)(r+s+d)}, \right.$$
$$\left. \frac{(1-\pi)(a+s)(s+d)(r+a+s) + \pi s(a+s+d)(r+s+d)}{\pi s(a+s+d)(r+s+d)}\right]$$

This completes the case and the proof. ■

Proof of Lemma 3 We prove this result by construction and observe that (2.39) is the relevant condition for obtaining the separating symmetric best response we seek. We will proceed in steps. First we will simplify the constraint set. Then we will characterize the solution to (2.25) with a reduced constraint set. Next we will solve for the income offers and finally we will verify that the dropped constraints are satisfied at the optimum.

Step 1 Guess that the relevant constraints for a firm facing problem (2.25) are constraints (2.26‴), (2.30‴), and a monotonicity constraint that $L_l \leq L_h$. The maximization problem seeks to minimize the incomes offered and it must be that (2.26‴) and (2.30‴) bind at the optimum. The binding of (2.26‴) with the monotonicity constraint implies that (2.29‴) is satisfied. The binding constraints can be solved to obtain

$$y_h = \frac{r+s+d}{d}\left(\theta_h L_h + (\theta_l - \theta_h)L_l\right) + \frac{a}{r+a+s}(\widetilde{y}_l - \theta_l \widetilde{L}_l) \tag{B.36}$$

$$y_l = \frac{r+s+d}{d}\theta_l L_l + \frac{r+s}{d}\theta_h(L_h - L_l) + \frac{a}{r+a+s}(\widetilde{y}_l - \theta_l \widetilde{L}_l) \tag{B.37}$$

The monotonicity constraint with (B.37) imply that (2.31‴) is also satisfied. We will continue to solve the problem and will check that (2.27‴) and (2.28‴) are indeed satisfied at the optimum.

Step 2 The problem for the firm with the new constraint set is to solve

$$\begin{aligned} \max_{L_l, L_h, n} \; & f(n(\pi L_h + (1-\pi)L_l)) \\ & - n\left[L_h\theta_h\left(\frac{r+s}{d} + \pi\right) + L_l\left(\frac{r+s}{d}(\theta_l - \theta_h) + \theta_l - \pi\theta_h\right)\right] \qquad (B.38) \\ & - n\frac{a}{r+a+s}(\widetilde{y}_l - \theta_l \widetilde{L}_l) - c \end{aligned}$$

$$s.t. \quad L_h \leq H \tag{B.39}$$

$$L_l \leq L_h \tag{B.40}$$

$$h \leq L_l \tag{B.41}$$

Letting the Lagrange multipliers on (B.39) - (B.41) be λ_1, λ_2, and λ_3, respectively,

the first order conditions (which are necessary and sufficient) for this problem are

$$[L_h]: f'(\cdot)n\pi = n\theta_h\left(\frac{r+s}{d}+\pi\right)+\lambda_1-\lambda_2 \tag{B.42}$$

$$[L_l]: f'(\cdot)n(1-\pi) = n\left(\frac{r+s}{d}(\theta_l-\theta_h)+\theta_l-\pi\theta_h\right)+\lambda_2-\lambda_3 \tag{B.43}$$

$$\begin{aligned}[n]: f'(\cdot)(\pi L_h+(1-\pi)L_l) &= L_h\theta_h\left(\frac{r+s}{d}+\pi\right)\\ &+L_l\left(\frac{r+s}{d}(\theta_l-\theta_h)+\theta_l-\pi\theta_h\right)+\frac{a}{r+a+s}(\widetilde{y}_l-\theta_l\widetilde{L}_l)\end{aligned} \tag{B.44}$$

The first order conditions imply that

$$L_h(\lambda_1-\lambda_2)+L_l(\lambda_2-\lambda_3)=n\frac{a}{r+a+s}(\widetilde{y}_l-\theta_l\widetilde{L}_l) \tag{B.45}$$

Note that $\lambda_1>0$. Otherwise, from (B.45) we have $\lambda_2(L_l-L_h)-\lambda_3 L_l>0$, which is a contradiction since $\lambda_i\geq 0$ for all i. Thus, with $\lambda_1>0$, we have that $L_h=H$. Hence, $\lambda_1-\lambda_2=n\frac{a}{H(r+a+s)}(\widetilde{y}_l-\theta_l\widetilde{L}_l)-\frac{L_l}{H}(\lambda_2-\lambda_3)$. Plugging this expression for $\lambda_1-\lambda_2$ into (B.42) and dividing (B.42) by (B.43), we have

$$\frac{\pi}{1-\pi}=\frac{\theta_h\left(\frac{r+s}{d}+\pi\right)+\frac{a}{H(r+a+s)}(\widetilde{y}_l-\theta_l\widetilde{L}_l)-\frac{L_l}{nH}(\lambda_2-\lambda_3)}{\frac{r+s}{d}(\theta_l-\theta_h)+\theta_l-\pi\theta_h+\frac{\lambda_2-\lambda_3}{n}} \tag{B.46}$$

As in the proof of Lemma 1, in order for (B.46) to be satisfied, the ordering of the inequality

$$\frac{\pi}{1-\pi}\gtreqless\frac{\theta_h\left(\frac{r+s}{d}+\pi\right)+\frac{a}{H(r+a+s)}(\widetilde{y}_l-\theta_l\widetilde{L}_l)}{\frac{r+s}{d}(\theta_l-\theta_h)+\theta_l-\pi\theta_h} \tag{B.47}$$

implies a sign of $\lambda_2-\lambda_3$, which yields the optimal solution of L_l as depicted in the statement of the lemma.

Step 3 To find a symmetric best response, let $L_h=\widetilde{L}_h=H$, $L_l=\widetilde{L}_l=L_l^*$, and $y_i=\widetilde{y}_i$ for all i in (B.36) and (B.37). This yields the desired characterization of the contracts in the lemma. Note that we can also substitute the expression for

$y_l - \theta_l L_l = \frac{r+a+s}{d}(\theta_h H + (\theta_l - \theta_h)L_l^*)$ into (B.47) to obtain (2.39), (2.40), and (2.41).

Step 4 Finally, it remains to verify that the constraints (2.27‴) and (2.28‴) are satisfied at the optimum. Employing the characterization of the labor contracts in the lemma, constraint (2.27‴) is equivalent to

$$\frac{a}{d}\theta_h H + \frac{r+a+s+d}{d}L_l^*(\theta_l - \theta_h) \geq \frac{a}{r+a+s}(y_h^* - \theta_h L_h^*)$$
$$= \frac{a}{r+a+s}\left(\frac{r+a+s}{d}\theta_h H + \frac{r+a+s+d}{d}L_l^*(\theta_l - \theta_h)\right)$$
$$\iff$$
$$1 \geq \frac{a}{r+a+s}$$

which is true. To verify constraint (2.28), note that it is equivalent to

$$\left(r+s+d - \frac{ad}{r+a+s}\right)\left(\frac{r+a+s}{d}\theta_h H + \frac{r+a+s+d}{d}L_l^*(\theta_l - \theta_h)\right)$$
$$\geq (r+s)\left(\frac{r+a+s+d}{d}\theta_l L_l^* + \frac{r+a+s}{d}\theta_h(H - L_l^*)\right)$$
$$\iff$$
$$\frac{r+a+s+d}{r+a+s}\left(\frac{r+a+s}{d}\theta_h H + \frac{r+a+s+d}{d}L_l^*(\theta_l - \theta_h)\right)$$
$$\geq \frac{r+a+s+d}{d}\theta_l L_l^* + \frac{r+a+s}{d}\theta_h(H - L_l^*)$$
$$\iff$$
$$\frac{r+a+s+d}{d}(\theta_h(H - L_l^*) + \theta_l L_l^*) + \frac{r+a+s+d}{r+a+s}L_l^*(\theta_l - \theta_h)$$
$$\geq \frac{r+a+s+d}{d}\theta_l L_l^* + \frac{r+a+s}{d}\theta_h(H - L_l^*)$$
$$\iff$$
$$\theta_h(H - L_l^*) + \frac{r+a+s+d}{r+a+s}L_l^*(\theta_l - \theta_h) \geq 0$$

which is true. This completes the proof. ■

Proof of Lemma 4 To simplify the problem in (2.43) we assume that the only relevant constraints for the problem are (2.28‴), (2.45′), and $L_h \leq L_l$. Note that (2.28‴) and (2.45′) must bind at the optimum. The monotonicity constraint and (2.45′) binding imply that (2.26‴) is satisfied. The two binding constraints imply that

$$y_h = y_l = \frac{r+s+d}{d}\theta_h L_h + \frac{a}{r+a+s}(y_h^* - \theta_h L_h^*) \tag{B.48}$$

which implies that (2.27‴) is satisfied (and in fact binds). Ignoring constraints (2.44′) and (2.46′), the firm problem now is

$$\max_{h \leq L_h \leq L_l \leq H, n} f(n\psi L_h) - n\left(\frac{r+s+d}{d}\theta_h L_h + \frac{a}{r+a+s}(y_h^* - \theta_h L_h^*)\right) - c$$

The firm will clearly set L_h at its maximum level, and so as to slacken the monotonicity constraint, the optimal solution will be one in which $L_h = L_l = H$ (note that L_l does not enter the objective function). The desired expression for $y_h = y_l$ is obtained by plugging the expression for $y_h^* - \theta_h L_h^*$ from Lemma 3 into (B.48). It remains to verify that (2.44′) and (2.46′) are satisfied at this optimum. Note that since the optimum is a pooling contract, (2.44′) and (2.46′) are equivalent constraints on the firm. Thus, it is sufficient to check that only one of them is satisfied. By plugging the expressions for y_l, y_l^*, L_l, and L_l^* into (2.46′), we see that constraint (2.46′) is equivalent to

$$\begin{aligned} \frac{r+s+d}{d}\theta_h H + \frac{a}{d}\left(\theta_h H + \frac{r+a+s+d}{r+a+s}(\theta_l - \theta_h)L_l^*\right) & \\ \leq \frac{r+s+d}{d}\theta_l H + \frac{a}{d}\left(\theta_h H + (\theta_l - \theta_h)L_l^*\right) & \\ \Longleftrightarrow & \\ \frac{a}{d}(\theta_l - \theta_h)L_l^*\left(\frac{r+a+s+d}{r+a+s} - 1\right) \leq \frac{r+s+d}{d}(\theta_l - \theta_h)H & \\ \Longleftrightarrow & \\ \frac{a}{r+a+s}L_l^* \leq \frac{r+s+d}{d}H & \end{aligned}$$

which is true. This completes the proof. ■

Proof of Lemma 5

For the high types we have that

$$\begin{aligned}
y_h - \theta_h L_h &\geq \frac{r+s}{d}\theta_h L_h + \frac{a}{r+a+s}(y_h^* - \theta_h L_h^*) \text{ by } (2.27''') \\
&\geq \frac{a}{r+a+s}(y_h^* - \theta_h L_h^*) \\
&= rU(\theta_h) \text{ by } (2.38) \\
\Leftrightarrow V_N\left(\frac{y_h}{L_h}, L_h; \theta_h\right) &\geq U(\theta_h)
\end{aligned}$$

And for the low types,

$$\begin{aligned}
V_S\left(\frac{y_l}{L_l}, L_l; \theta_l\right) &\geq U(\theta_l) \\
\Leftrightarrow y_l &\geq rU(\theta_l) \\
&= \frac{a}{r+a+s}(y_l^* - \theta_l L_l^*) \text{ by } (2.38) \\
&= \frac{a}{d}\theta_h H + \frac{a}{d}(\theta_l - \theta_h)L_l^* \text{ by Lemma 3} \\
\Leftrightarrow \frac{r+s+d}{d}\theta_h H &+ \frac{a}{r+a+s}(\theta_l - \theta_h)L_l^* \geq 0 \text{ by Lemma 4}
\end{aligned}$$

which is true. ■

Proof of Proposition 5 To prove this result, we will begin by considering the more general case of a potential non-shirking pure strategy separating symmetric Nash equilibrium in which $L_l^* = h$. From the work in Lemmas 3, 4, and 5, for the desired equilibrium to exist, there must be no profitable deviation to a pooling contract which allows for shirking from the low type:

$$
\begin{aligned}
&\max_n f(n(\pi H + (1-\pi)h)) \\
&\quad - n\left[H\theta_h\left(\frac{r+s}{d}+\pi\right) + h\left(\frac{r+s}{d}(\theta_l-\theta_h)+\theta_l-\pi\theta_h\right)\right] \\
&\quad - n\frac{a}{d}(\theta_h H + (\theta_l-\theta_h)h) - c \\
&\geq \\
&\max_n f(n\psi H) - n\left[\frac{r+s+d}{d}\theta_h H + \frac{a}{d}\left(\theta_h H + \frac{r+a+s+d}{r+a+s}(\theta_l-\theta_h)h\right)\right] - c
\end{aligned}
$$

This is equivalent to comparing the cost of effective labor,

$$F(\beta) \geq G(\beta)$$

where we define

$$
\begin{aligned}
\beta &\equiv \frac{h}{H} \in [0,1] \\
F(\beta) &\equiv \frac{\frac{r+a+s+d}{d}\theta_h + \frac{a}{d}\frac{r+a+s+d}{r+a+s}(\theta_l-\theta_h)\beta}{\psi} \\
G(\beta) &\equiv \frac{\frac{r+a+s}{d}(\theta_h + (\theta_l-\theta_h)\beta) + \pi\theta_h + \beta(\theta_l - \pi\theta_h)}{\pi + (1-\pi)\beta}
\end{aligned}
$$

In addition, for a fully separating equilibrium we must have the constraint on parameters given by (2.39) (reproduced here using the β transformation of hours):

$$\frac{\pi}{1-\pi} \geq \frac{\theta_h\left(\frac{r+s}{d}+\pi\right) + \frac{a}{d}\left(\theta_h + (\theta_l-\theta_h)\beta\right)}{\frac{r+s}{d}(\theta_l-\theta_h)+\theta_l-\pi\theta_h} \tag{2.39$'$}$$

This constraint places an upper bound on β. In particular, (2.39$'$) is equivalent to

$$\beta \leq \overline{\beta} \equiv \frac{\frac{d}{a}\left(\frac{\pi}{1-\pi}\frac{r+s+d}{d}\theta_l - \frac{1}{1-\pi}\left(\frac{r+s}{d}+\pi\right)\theta_h\right) - \theta_h}{\theta_l - \theta_h} \tag{2.39$''$}$$

For a separating equilibrium to exist it must be that $\overline{\beta} \geq 0$. I claim that in fact, $G(\beta) > F(\beta) \forall \beta \in [0,\overline{\beta}] \cap [0,1)$, making a deviation always profitable for all parameterizations of a potential separating equilibrium model.

To show this claim, it will be helpful to remember that from the analysis in Lemma 3 regarding the solution to problem (2.25), if all other firms are offering the fully separating contracts, $\{(y_i^*, L_i^*)\}$ and $\beta > \overline{\beta}$, then a single firm will prefer to offer a pooling contract which induces effort from everyone. In particular, the firm will offer everyone H hours of work with an income of $y = \frac{r+s+d}{d}\theta_l H + \frac{a}{d}\left((\theta_l - \theta_h)h + \theta_h H\right)$.[1] In summary, for $\beta > \overline{\beta}$,

$$
\begin{aligned}
&\max_n f(nH) - n\left[\frac{r+s+d}{d}\theta_l H + \frac{a}{d}\left((\theta_l - \theta_h)h + \theta_h H\right)\right] - c \\
&\qquad > \\
&\max_n f(n(\pi H + (1-\pi)h)) \\
&\qquad - n\left[H\theta_h\left(\frac{r+s}{d} + \pi\right) + h\left(\frac{r+s}{d}(\theta_l - \theta_h) + \theta_l - \pi\theta_h\right)\right] \\
&\qquad - n\frac{a}{d}(\theta_h H + (\theta_l - \theta_h)h) - c
\end{aligned}
$$

We can equivalently write this as

$$G(\beta) > P(\beta) \text{ for } \beta > \overline{\beta} \tag{B.49}$$

where

$$P(\beta) \equiv \frac{r+s+d}{d}\theta_l + \frac{a}{d}\left[(\theta_l - \theta_h)\beta + \theta_h\right]$$

Moreover, for $\beta < \overline{\beta}$, the inequality in (B.49) is reversed and $G(\overline{\beta}) = P(\overline{\beta})$. This, with $\overline{\beta} \geq 0$ imply that $P(0) \geq G(0)$.

[1]This expression for income is computed from plugging $L_l = L_h = H$ into (B.36) and (B.37). It is also straightforward to verify that constraints (2.27‴) and (2.28‴) are satisfied and were therefore legitimately ignored in the maximization problem.

In addition, observe that:

$$G(0) > F(0)$$

$$\Longleftrightarrow$$

$$\frac{\left(\frac{r+a+s}{d} + \pi\right)\theta_h}{\pi} > \frac{\left(\frac{r+a+s}{d} + 1\right)\theta_h}{\psi}$$

$$\Longleftrightarrow$$

$$\frac{\pi(s+d)}{s+\pi d}\left(\frac{r+a+s}{d} + \pi\right) > \pi\left(\frac{r+a+s}{d} + 1\right)$$

$$\Longleftrightarrow$$

$$1 < \frac{r+a+s}{d}\left(\frac{s+d}{s+\pi d} - 1\right) + \frac{\pi(s+d)}{s+\pi d} = \frac{(1-\pi)(r+a)}{s+\pi d} + 1$$

which is true.

It is also helpful to characterize how the functions, $F(\cdot), G(\cdot)$, and $P(\cdot)$ vary with β in order to determine their ordering. Both $F(\cdot)$ and $P(\cdot)$ are linearly increasing functions of β, with $F'(\beta) > P'(\beta)$. In addition,

$$G'(\beta) = \frac{(\theta_l - \theta_h)\pi + \frac{r+a+s}{d}(\pi\theta_l - \theta_h)}{(\pi + (1-\pi)\beta)^2}$$

$$G''(\beta) = \frac{-2(1-\pi)}{\pi + (1-\pi)\beta} G'(\beta)$$

In order to sign these derivatives, note that

$$\overline{\beta} \geq 0 \Longleftrightarrow \frac{\theta_h}{\theta_l} \leq \frac{\pi(r+s+d)}{r+s+\pi d + (1-\pi)a} \tag{B.50}$$

which is necessary for existence of a separating equilibrium. In addition, one can show with algebra that

$$\frac{\pi(r+s+d)}{r+s+\pi d + (1-\pi)a} < \frac{\pi(r+a+s+d)}{r+a+s+\pi d} \tag{B.51}$$

Finally, note that

$$G'(\beta) > 0 \iff \frac{\theta_h}{\theta_l} < \frac{\pi(r+a+s+d)}{r+a+s+\pi d} \tag{B.52}$$

Taking (B.50) - (B.52) together, it follows that $G'(\beta) > 0$, and since $G''(\beta)$ is of opposite parity of $G'(\beta)$, we also know that $G(\cdot)$ is concave.

We are now ready to show that indeed $G(\beta) > F(\beta) \forall \beta \in [0, \overline{\beta}] \cap [0,1)$ as claimed. We proceed in cases.

Case 1: $\overline{\beta} \geq 1$.

Since $G(0) > F(0)$, F is linearly increasing, and G is increasing and concave, if $G(1) > F(1)$ it follows that $G(\beta) > F(\beta) \forall \beta \in [0,1)$ as claimed. Suppose on the contrary that $F(1) \geq G(1)$. When $\beta = 1$ there is no hours choice and the problem reduces to the first analysis in the chapter in which the no deviation condition is equivalent to the inequality in (2.13), reprinted:

$$\frac{\theta_h}{\theta_l} \geq \frac{\pi(s+d)}{s+\pi d} - \frac{as(1-\pi)}{(s+\pi d)(r+s)} \tag{2.13}$$

In addition, the assumption that $\overline{\beta} \geq 1$ is equivalent to

$$\frac{\theta_h}{\theta_l} \leq \frac{\pi(r+s+d) - a(1-\pi)}{r+s+\pi d} \tag{B.53}$$

Thus, combining (B.53) and (2.13), we have

$$\frac{\pi(r+s+d) - a(1-\pi)}{r+s+\pi d} \geq \frac{\pi(s+d)}{s+\pi d}\frac{r+a+s}{r+s} - \frac{a}{r+s}$$

$$\iff$$

$$0 \geq r+s+a$$

which is clearly a contradiction. Hence, $G(1) > F(1)$, delivering the desired result.

Case 2: $\overline{\beta} < 1$

Subcase a: $P(1) \geq F(1)$

Since $G(1) > P(1)$ by (B.49) and $\overline{\beta} < 1$, it follows from the assumption of

$P(1) \geq F(1)$ that $G(1) > F(1)$. For the same reasoning as in Case 1, this guarantees that $G(\beta) > F(\beta) \forall \beta \in [0, 1)$.

Subcase b: $P(1) < F(1)$

Since $\overline{\beta} < 1$, it is sufficient to show that $G(\beta) > F(\beta) \forall \beta \in [0, \overline{\beta}]$. This will be true, as in the previous cases, if $G(\overline{\beta}) > F(\overline{\beta})$. A sufficient condition to guarantee this inequality is for $\overline{\beta} < \hat{\beta}$ where $\hat{\beta}$ is defined by $F(\hat{\beta}) = P(\hat{\beta})$. To see this, note that both $G(\cdot)$ and $F(\cdot)$ must each intersect $P(\cdot)$ once from below at $\overline{\beta}$ and $\hat{\beta}$, respectively. Thus, if $\overline{\beta} < \hat{\beta}$, $G(\cdot)$ intersects $P(\cdot)$ at a point in which $P(\cdot) > F(\cdot)$, implying transitively that $G(\overline{\beta}) > F(\overline{\beta})$, as desired. The figure below illustrates this argument graphically.

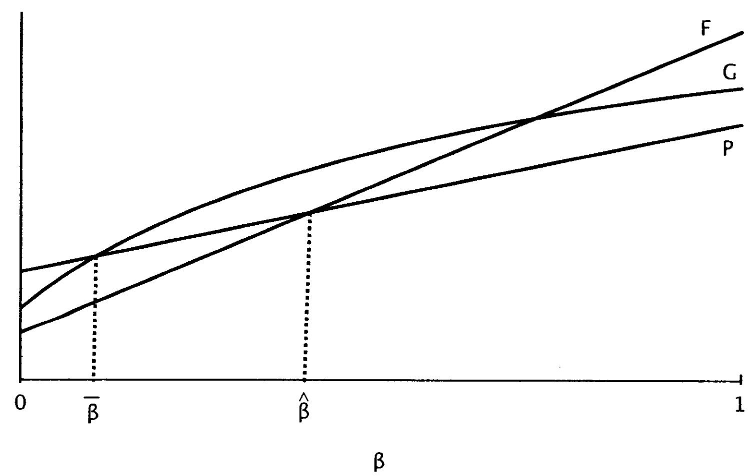

Figure B-2: $\overline{\beta} < \hat{\beta}$

To prove that $\overline{\beta} < \hat{\beta}$, we must first determine $\hat{\beta}$. By definition,

$$F(\hat{\beta}) = P(\hat{\beta}) \iff \hat{\beta} = \frac{\psi\left(\frac{r+s+d}{d}\theta_l + \frac{a}{d}\theta_h\right) - \frac{r+a+s+d}{d}\theta_h}{(\theta_l - \theta_h)\left(\frac{a}{d}\frac{r+a+s+d}{r+a+s} - \psi\frac{a}{d}\right)}$$

Thus, we want to show that

$$
\begin{aligned}
& \hat{\beta} > \overline{\beta} \\
\iff & \frac{\frac{d}{a}\psi\left(\frac{r+s+d}{d}\theta_l + \frac{a}{d}\theta_h\right) - \frac{r+a+s+d}{a}\theta_h}{\left(\frac{r+a+s+d}{r+a+s} - \psi\right)} \\
& \qquad > \frac{d}{a}\left(\frac{\pi}{1-\pi}\frac{r+s+d}{d}\theta_l - \frac{1}{1-\pi}\left(\frac{r+s}{d}+\pi\right)\theta_h\right) - \theta_h \qquad \text{(B.54)} \\
\iff & \theta_l(r+s+d)((1-\pi)(r+a)-\pi(s+d)) \\
& \qquad > \theta_h(r(1-\pi)(r+a) - r\pi(s+d) - (s+d\pi)(s+d)) \qquad \text{(B.55)}
\end{aligned}
$$

where the last equivalence follows only after lengthy algebraic manipulation. To show that (B.55) holds, we consider three cases in turn, depending on the sign of each side of (B.55).

Case i: RHS ≥ 0

Observe that

$$
\begin{aligned}
& (r+s+d)((1-\pi)(r+a)-\pi(s+d)) \\
& \qquad > r(1-\pi)(r+a) - r\pi(s+d) - (s+d\pi)(s+d) \\
& \iff (1-\pi)(r+a+s) > 0
\end{aligned}
$$

which is true. This observation, the assumption that the right hand side is non-negative, and $\theta_l > \theta_h$ imply that (B.55) is satisfied.

Case ii: RHS < 0 *and LHS* ≥ 0

The inequality in (B.55) obviously holds in this case.

Case iii: RHS < 0 *and LHS* < 0

We proceed by contradiction. Suppose (B.55) does not hold, i.e.

$$
\begin{aligned}
& \theta_l(r+s+d)(\pi(s+d)-(1-\pi)(r+a)) \\
& \qquad \geq \theta_h(r\pi(s+d) - r(1-\pi)(r+a) + (s+d\pi)(s+d)) \qquad \text{(B.56)}
\end{aligned}
$$

In addition, the assumption that $\overline{\beta} < 1$ is equivalent to $\theta_l < \theta_h \frac{r+s+d\pi}{(r+s+d)\pi - a(1-\pi)}$. This combined with (B.56) implies that

$$
\begin{aligned}
&\frac{\theta_h(r+s+d\pi)(r+s+d)(\pi(s+d)-(1-\pi)(r+a))}{(r+s+d)\pi-a(1-\pi)} \\
&> \theta_h(r\pi(s+d)-r(1-\pi)(r+a)+(s+d\pi)(s+d)) \\
&\iff (1-\pi)(r+a+s+d) < 0
\end{aligned}
$$

where the last inequality follows from lengthy algebraic manipulation and is clearly a contradiction. Thus, (B.55) is satisifed. This completes the nonexistence proof for the fully separating case.

To see that a non-shirking pure strategy symmetric Nash equilibrium with $L_l^* \in (h, H)$ cannot exist either, note that for such an equilibrium, we need (2.41). In such a case, the firm is indifferent between all choices of L_l^* and in particular

$$
\begin{aligned}
&\max_n f(n(\pi H+(1-\pi)L_l^*)) \\
&\qquad - n\left[\frac{r+a+s}{d}(\theta_h H+(\theta_l-\theta_h)L_l^*)+\pi\theta_h H+(\theta_l-\pi\theta_h)L_l^*\right]-c \\
&= \\
&\max_n f(nH) - n\left[\frac{r+s+d}{d}\theta_l H+\frac{a}{d}\left(\theta_h H+(\theta_l-\theta_h)L_l^*\right)\right]-c
\end{aligned}
$$

where the cost in the second profit expression is simply found by plugging $L_l = L_h = H$ and $y_l^* - \theta_l L_l^*$ from Lemma 3 into (B.36).

In addition, we must have no profitable deviation by a single firm to allowing for shirking by the low type agents. As before, this requires that

$$
\begin{aligned}
&\max_n f(nH) - n\left[\frac{r+s+d}{d}\theta_l H+\frac{a}{d}\left(\theta_h H+(\theta_l-\theta_h)L_l^*\right)\right]-c \\
&\geq \\
&\max_n f(n\psi H) - n\left[\frac{r+s+d}{d}\theta_h H+\frac{a}{d}\left(\theta_h H+\frac{r+a+s+d}{r+a+s}(\theta_l-\theta_h)L_l^*\right)\right]-c
\end{aligned}
$$

which is equivalent to showing that

$$\frac{r+s+d}{d}\theta_h+\frac{a}{d}\left(\theta_h+\frac{r+a+s+d}{r+a+s}(\theta_l-\theta_h)\beta^*\right)$$
$$\geq \psi\left(\frac{r+s+d}{d}\theta_l+\frac{a}{d}(\theta_h+\beta^*(\theta_l-\theta_h))\right)$$

where $\beta^* \equiv \frac{L_l^*}{H}$. Rearranging this expression, we have that no profitable deviation is equivalent to

$$\beta^*(\theta_l-\theta_h) \geq \frac{\psi\left(\frac{r+s+d}{d}\theta_l+\frac{a}{d}\theta_h\right)-\frac{r+a+s+d}{d}\theta_h}{\frac{a}{d}\left(\frac{r+a+s+d}{r+a+s}-\psi\right)} \tag{B.57}$$

From (2.41) we have that

$$\beta^*(\theta_l-\theta_h)=\frac{d}{a}\left(\frac{\pi}{1-\pi}\frac{r+s+d}{d}\theta_l-\frac{1}{1-\pi}\left(\frac{r+s}{d}+\pi\right)\theta_h\right)-\theta_h \tag{B.58}$$

From (B.54), however, we know that (B.58) yields a reversed inequality in (B.57). Thus, there is a profitable deviation and nonexistence of a non-shirking pure strategy symmetric Nash equilbrium with partial separation.

This completes the proof of the proposition. ■

Bibliography

[1] Stefania Albanesi and Christopher Sleet. Dynamic optimal taxation with private information. *Review of Economic Studies*, 73:1–30, 2006.

[2] James Albrecht and Susan B. Vroman. Nash equilibrium efficiency wage distributions. *International Economic Review*, 39:183–203, 1998.

[3] Robin Boadway, Manuel Leite-Monteiro, Maurice Marchand, and Pierre Pestieau. Social insurance and redistribution. In S Cnossen and H. W. Sinn, editors, *Public Finance and Public Policy in the New Century*, pages 333–358. MIT Press, Cambridge, MA, 2003.

[4] Robin Boadway, Manuel Leite-Monteiro, Maurice Marchand, and Pierre Pestieau. . Social insurance and redistribution with moral hazard and adverse selection. *Scandinavian Journal of Economics*, 108:279–298, 2006.

[5] Ching-Huei Chang. Optimal taxation in an efficiency wage model. *Southern Economic Journal*, 62:428–439, 1995.

[6] Raj Chetty. Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. forthcoming, Annual Reviews of Economics, 2009.

[7] Raj Chetty and Emmanuel Saez. Optimal taxation and social insurance with endogenous private insurance. unpublished working paper: http://elsa.berkeley.edu/ saez/chetty-saez09socialins.pdf, 2009.

[8] Helmuth Cremer and Pierre Pestieau. Redistributive taxation and social insurance. *International Taxation and Public Finance*, 3:259–280, 1996.

[9] Julie Berry Cullen and Jonathan Gruber. Does unemployment insurance crowd out spousal labor supply? *Journal of Labor Economics*, 18:546–572, 2000.

[10] David Cutler and Jonathan Gruber. Does public insurance crowd out private insurance? *Quarterly Journal of Economics*, 111:391–430, 1996.

[11] David Cutler and Jonathan Gruber. The effect of medicaid expansions on public insurance, private insurance and redistribution. *American Economics Review*, 86:378–383, 1996.

[12] Peter Diamond, J. Helms, and James Mirrlees. Optimal taxation in a stochastic economy: a cobb-douglas example. *Journal of Public Economics*, 14:1–30, 1980.

[13] Jonathan Eaton and Harvey Rosen. Taxation, human capital, and uncertainty. *Journal of Public Economics*, 14:365–374, 1980.

[14] Amy Finkelstein. The interaction of partial public insurance programs and residual private insurance markets: evidence from the us medicare program. *Journal of Health Economics*, 23:1–24, 2004.

[15] Michael Golosov and Aleh Tsyvinski. Optimal taxation with endogenous insurance markets. *Quarterly Journal of Economics*, 121:487–534, 2007.

[16] Mikhail Golosov and Aleh Tsyvinski. Designing optimal disability insurance: A case for asset testing. *Journal of Political Economy*, 114:257–279, 2006.

[17] Jonathan Gruber and Kosali Simon. Crowd-out 10 years later: Have recent public insurance expansions crowded out private health insurance? *Journal of Health Economics*, 27:201–217, 2008.

[18] Veronica Guerrieri, Robert Shimer, and Randall Wright. Adverse selection in competitive search equilibrium. Econometrica, revise and resubmit, 2009.

[19] G.E. Johnson and P.R. G. Layard. *The Natural Rate of Unemployment: Explanation and Policy*, volume 2, pages 921–999. Elsevier Science Publishers, Amsterdam, 1986.

[20] Louis Kaplow. Incentives and government relief for risk. *Journal of Risk and Uncertainty*, 3:167–175, 1991.

[21] Narayana Kocherlakota. Zero expected wealth taxes: A mirrlees approach to dynamic optimal taxation. *Econometrica*, 73:1587–1621, 2005.

[22] Brandon Lehr. Correction: Albrecht and vroman's nonexistence proof of symmetric nash equilibrium with efficiency wages. MIT Thesis Chapter, 2010.

[23] Brandon Lehr. Efficiency wages with heterogeneous agents. MIT Thesis Chapter, 2010.

[24] James Mirrlees. An exploration in the theory of optimum income taxation. *Review of Economic Studies*, 38:175–208, 1971.

[25] James Mirrlees. Taxing uncertain incomes. *Oxford Economic Papers*, 42:34–45, 1990.

[26] Giuseppe Pisauro. The effect of taxes on labor in efficiency wage models. *Journal of Public Economics*, 46:329–345, 1991.

[27] Jean-Charles Rochet. Incentives, redistribution and social insurance. *The Geneva Papers on Risk and Insurance*, 16:143–165, 1991.

[28] C. Shapiro and J. Stiglitz. Equilibrium unemployment as a worker discipline device. *American Economic Review*, 74:433–444, 1984.

[29] J. Strand. Unemployment as a discipline device with heterogeneous labor. *American Economic Review*, 77:489–493, 1987.

[30] Hal Varian. Redistributive taxation as social insurance. *Journal of Public Economics*, 14:49–68, 1980.

[31] Janet L. Yellen. Efficiency wage models of unemployment. *American Economic Review*, 74:200–205, 1984.