Review of Economic Studies (2011) **78**, 394–428 doi: 10.1093/restud/rdq010

General Equilibrium Returns to Human and Investment Capital under Moral Hazard

CHRISTINE A. PARLOUR and JOHAN WALDEN
Haas School of Business, University of California, Berkeley

First version received May 2008*; final version accepted May* 2010 (*Eds.*)

We present a tractable general equilibrium model with multiple sectors in which firms offer workers incentive contracts and simultaneously raise capital in stock markets. Workers optimally invest in the stock market and at the same time hedge labour income risk. Firms rationally take agents' portfolio decisions into account. In equilibrium, the cost of capital of each sector is endogenous. The distortion induced by moral hazard generates counterintuitive effects on the real economy. For example, the value of labour market participation may be higher under moral hazard than under first best, further a positive productivity shock may decrease welfare in the moral hazard economy. In addition, our model generates predictions on the effects of moral hazard on asset markets. For example, in the presence of moral hazard, the capital asset pricing model fails because firms, by choosing optimal incentive contracts, transfer risk both through wages and through the stock market. This leads to several cross-sectional asset pricing "anomalies", such as size and value effects. As we characterize optimal contracts, we can also present empirical predictions relating workers' compensation, firm productivity, firm size, and financial market abnormal returns.

JEL Codes: G10, G12

1. INTRODUCTION

In the U.S., labour accounts for about two-thirds of national income and there is heterogeneity in the source of this income. Each worker is typically tied to one industry, be it a hedge fund in New York or almond production in California, and all workers differ in the marginal productivity of their labour. Further, households with labour income are three times more likely to own financial assets than those without.[1] Given that workers supply human and investment capital to firms, how do firms optimally compensate workers, how do risk-averse investors value firms, and what effect does this have on real production and asset prices?

To answer these questions, we present a general equilibrium model with a continuum of workers and many firms that are organized into different sectors. In the first period, workers endowed with wealth accept employment contracts offered by firms and their effort is used as an input into production. Firms in different sectors have different labour productivities and therefore, in equilibrium, elicit different effort levels from their employees. The only source of heterogeneity across investors is the difference in employment contracts. To motivate these, we follow Holmström and Milgrom (1987) and assume that workers control the drift of a revenue

1. According to the 2004 survey of consumer finances, approximately 60% of households with labour income own financial assets, while approximately 20% of households with no labour income own financial assets.

process by exerting costly effort. Firm revenue depends both on systematic (industry wide) and on idiosyncratic risk, and workers' contracts contain both. Workers are prohibited from shorting the firm in which they work, however they can trade in an industry portfolio that is perfectly correlated with industry-specific risk. The idiosyncratic risk ensures that the workers supply effort, but they lay off any systematic risk in wages by trading in the stock market.

At the firm level, the endogenous variables are the wage contract (and attendant employee effort) and firms' initial investment level (which is just firm size). Workers choose effort levels and an optimal hedging portfolio; these differ across industries. In general equilibrium, all markets clear and firms make zero economic profits. This pins down the real variables in the economy: firm and industry size and effort levels in each industry. In addition, we determine the equilibrium value of human capital as all workers must be indifferent between their current employment and moving to another industry. The equilibrium financial variables are the prices of the indexes in industry-specific risk and the cost of capital for each firm.

We show that the equilibrium effects of moral hazard are non-trivial and that intuition derived from partial equilibrium analysis often fails. Although the total welfare is lower in an economy with moral hazard, the effects on individual markets are ambiguous. For example, even though moral hazard makes it more difficult for firms to contract with workers, we show that agents can actually be better off participating in the labour market of an economy with moral hazard than in the first-best economy. In other words, the certainty equivalent of labour market participation can be higher under moral hazard than the first best. This can happen if a change in productivity in a sector leads agents to change their hedging demand and increase overall investment. Similarly, we show that the certainty equivalent of asset market participation may be higher under moral hazard than the first best. This result holds if moral hazard endogenously leads to industry sectors that are more "balanced" in size.

Positive productivity shocks to labour can decrease welfare in the moral hazard economy. This also works through the investment channel: a productivity shock increases welfare if it increases the equilibrium marginal product of investment, this can only be true if new capital flows into the industry as a result of the shock or if entry costs are sufficiently low. If they are not, then one can find examples of economies in which productivity shocks decrease welfare in the moral hazard economy but would increase welfare if the shock occurred in a first-best economy. We conclude that any partial equilibrium analysis of the effect of moral hazard in a specific market may be misleading.

One result holds under general conditions in our economy: increased moral hazard in a sector unambiguously decreases the size of that sector (although not the sizes of individual firms in that sector). Therefore, the distribution of risks faced by investors in the economy is affected by the moral hazard problem. This is important because the economy-wide distribution of risk determines risk-averse agents' valuation for any specific industry risk and therefore each firm's cost of capital. The result may also be of empirical interest since it provides a link between (observable) industry size and (unobservable) moral hazard.

We also study the asset pricing implications of moral hazard. If there are no frictions in the labour market, then a simple market-based capital asset pricing model (CAPM) holds. This is not the case in the presence of moral hazard. To provide incentives, the firm pays the worker a proportion of its revenue stream and so wages contain both systematic and idiosyncratic risk. *A priori*, one might suppose that this would lead to significant welfare costs since the worker is risk averse. However, since the worker can trade, he will *ex ante* hedge away the systematic risk in the market, whereas because of trading restrictions, he bears the idiosyncratic risk. The non-separation is therefore irrelevant to the worker.

In general equilibrium, higher-productivity industries are larger and therefore have higher returns. However, in order to induce high effort these industries pay out a higher proportion of the wage bill in profit sharing. Therefore, the stocks of these industries seem less risky than they really are, and so relative to the stock market portfolio, the stock returns are too high. These effects can only be understood in a general equilibrium framework since the link between firm characteristics, incentive contracts, systematic risk, and industry cost of capital is quite arbitrary in a partial equilibrium setting. Through our assumptions on the production function, we demonstrate that in general equilibrium more productive firms are smaller. This allows us to link characteristics of the real economy to observed abnormal returns; thus, the model yields a "size effect" (Banz, 1981; Fama and French, 1993). We also generate a "value effect" in our model. Given that fixed wage compensation is more front loaded than variable compensation (bonuses can only be paid after outcomes are observed), low-productivity firms will have more front loaded wage compensation and thereby a lower book-to-market ratio. But these are exactly the firms that have lower expected returns.

As we price risk, we can also analyse firms' endogenous cost of capital. The standard asset pricing intuition is that if a sector is larger, in equilibrium its expected return must be higher to compensate (risk-averse) investors (see, *e.g.* Bansal, Fang and Yaron, 2006) may not hold when wage contracts are endogenous. Indeed, we show that this intuition may fail in the moral hazard economy. Intuitively, the presence of moral hazard shifts the channel through which risk is paid from the capital market to the labour market, which reverses the standard intuition.

We offer new predictions on the relationship between the characteristics of the wage bill and abnormal returns. Indeed, we provide explicit predictions on the relationship between compensation (either wage or profit sharing) and asset market returns. We also provide predictions on the relationship between agents' labour productivity and portfolio investments.

We interpret our asset pricing results in light of the Roll's (1977) critique of the CAPM. Roll's critique, that the only test of the CAPM is whether the market portfolio is efficient, follows from the fact that the true market portfolio is unobservable. In our model, risky labour income is the main source of the discrepancy between the observed and true market portfolio and drives the failure of the CAPM to price assets correctly. The consumption CAPM (Breeden, 1979), on the other hand, does hold in our model. In the context of these models, our analysis suggests that a supply-side proxy for the true market portfolio is given by stock returns together with firms' risky labour expenses. Empirically, while we do not frequently observe labour contracts, we do observe the returns to human capital. These reflect the risk paid out to workers in the form of wages. An empirical literature has tried to proxy for the investments that agents make through their human capital. For example, Jagannathan and Wang (1996) find that a labour factor improves predictability of expected returns.

Following the insights of Telmer (1993) and Heaton and Lucas (1996), idiosyncratic labour income and incomplete markets do not seem to fully resolve asset pricing anomalies in the time series. We take a different approach, namely, that labour income is related to the real economy and therefore might help with cross-sectional predictions. Specifically, in our model, labour income is tied to a particular sector of the economy and therefore generates a particular hedging demand that affects firms' cost of capital. A literature has developed to analyse the effect of labour income in explaining cross-sectional asset prices. Most recently, Santos and Veronesi (2006) demonstrate how including stochastic labour income in a representative agent economy generates return predictability. They find that the labour-to-consumption ratio is predictor of long-run returns. Danthine and Donaldson (2002) demonstrate in a dynamic model that the implicit leverage implied by wage payments combined with uninsurable labour income risk generates realistic equity premia. Our focus is different since we focus on the close cross-sectional link between wage compensation and returns in capital markets. In this respect, we are closer to Bodie, Merton and Samuelson (1992), who assume perfect correlation between human capital and stock return in a one asset portfolio choice model, and also to Qin

(2002) who introduces a similar model. However, compared with these two papers, our analysis goes further by endogenizing firms' labour compensation decisions and the industry cost of capital.

Acharya and Bisin (2009) present a model in which managers, who are prevented from trading in stocks correlated with their industry, try to pick technologies so that their eventual labor income is well diversified. They too use a constant absolute risk aversion (CARA) normal framework, however the nature of the moral hazard (their managers pick the correlation of the project with the market portfolio) is different than ours. Further, in their incomplete markets setting the objective function of the firm is not defined. In our economy, all agents agree on economic profit maximization or equivalently on the discount rate that the firm should use for systematic risk. Ou-Yang (2005) presents an equilibrium model of asset pricing and moral hazard. Our work is similar in that we exploit the tractability of the CARA-normal framework pioneered by Holmström and Milgrom (1987). A key difference between our frameworks is that in our model, workers are also the investors and can trade on any systematic risk in their compensation package. This ensures that the objective function of the firm is well defined (*i.e.* all investors agree on profit maximization). We also fully endogenize agents' participation wages.

These differences also distinguish our work from Zame (2007) who considers a general equilibrium model in which firms, firm organization, and the prices of inputs and outputs for all the consumption goods are endogenous. In his model, state contingent profit-sharing plans are part of the description of the firm, the prices of which are determined in equilibrium. In other words, intra-firm transfers replace an external asset market. By contrast, to distinguish between wage payments and asset market returns, we restrict attention to the particular utility function for which the objective function of the firm is well defined, *i.e.* all shareholders agree on a "net present value" investment rule. Further, by explicitly allowing all agents to trade in securities markets we can relate asset market returns to our endogenous wage contract.

2. MODEL

Consider the following two-date economy populated by firms and workers. At time $t=1$, workers sign employment contracts with firms and then trade in financial markets. Simultaneously, firms raise money in the financial markets and make investment decisions. Then the production phase begins: workers continuously exert effort between time $t=1$ and $t=2$, which generates firm cash flows. At $t=2$, workers are paid and then all firms are dissolved. The cumulative cash flows, net of wages, are paid out as a liquidating dividend. (Throughout the paper, we adhere to the convention that a boldface letter presents a vector, the superscript T denotes a transpose, and the operator $()_i$ selects the i-th element of a vector. For an arbitrary vector, $\boldsymbol{a}$, we use the notation $\mathrm{diag}(\boldsymbol{a})$ to denote the diagonal matrix with $\boldsymbol{a}$ on its diagonal.)

The economy is populated by a mass $M>0$ of *ex ante* identical agents indexed by m. Each agent has a CARA utility function

$$U_m = U(W_m, e_m) = -e^{-\rho\left(W_m - \frac{k\int_1^2 e_m^2(t)dt}{2}\right)}, \tag{1}$$

over $t=2$ wealth denoted by W_m and effort, $e_m(t)$, expended continuously by the agent between $t=1$ and $t=2$. The cost per unit effort expended by a worker is $k>0$. The $t=2$ wealth of agent m, W_m, is made up of his accumulated wage income, ω_m, and the pay-off of his investment portfolio, which we describe below. There are no wealth effects with CARA utility, and so we assume that at $t=1$ all agents have identical initial wealth W.

Workers can find employment in any one of $n = 1, \ldots, N$ sectors (also called industries[2]) each of which comprises a continuum of identical firms indexed by ℓ. Worker m earns a wage because at any point in time his effort level $e_m(t)$ affects the drift of the instantaneous revenue generated by the firm while in operation. Suppose that firm ℓ in industry n made a capital investment of $I_{n,\ell}$, then the firm's revenue evolves as

$$d\tilde{R}_{n,\ell} = I_{n,\ell}(\alpha_n e_m(t)dt + d\tilde{\varepsilon}_{n,\ell} + d\tilde{x}_n), \tag{2}$$

where $\tilde{R}_{n,\ell}(1) = 0$. Here, $\tilde{x}_n(t)$ is a Brownian motion that is common to all firms in an industry (although different across industries), with unit drift ($E[d\tilde{x}_n] = dt$), and $\text{cov}(d\tilde{x}_n, d\tilde{x}_k) = \sigma_{n,k}dt$ is the constant instantaneous covariance between the shocks in two separate industries. The Brownian motion $d\tilde{\varepsilon}_{n,\ell}$ is a firm-specific (idiosyncratic) shock that is uncorrelated across firms in the industry and across industries, with zero drift and instantaneous variance $\text{cov}(d\tilde{\varepsilon}_{n,\ell}, d\tilde{\varepsilon}_{n,\ell}) = \sigma^2_{\varepsilon,n}dt$. We also define $\sigma^2_{x,n} = \sigma_{n,n}$. The parameter α_n is the productivity of labour in industry n. The productivity of workers differs across industries, and without loss of generality, we assume that $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_N$.

The effort that the agent exerts at time t, $e_m(t)$, is the optimal response to the wage packet that he has been promised. We shall focus on economies in which workers in the same industry (optimally) choose the same effort, so henceforth we use the n subscript for effort, $e_n(t)$, $n = 1, \ldots, N$.

There are two sources of noise in the revenue process, one idiosyncratic to the firm and one common to the whole industry. Neither the firm nor the agent can distinguish between the two sources. However, both the firm and the worker observe the full history $\{\tilde{R}(\tau)_{n,\ell} \mid \tau \leq t\}$ and both the compensation and the effort may be contingent on this process. Following Holmström and Milgrom (1987), we will show that it is optimal for the firm management to convey a set of instructions to the worker and promise a compensation package at time $t = 2$ that is contingent on the realized path of the revenue process up to time $t = 2$, $\tilde{R}_{n,\ell}(2)$. This also implies that once we have shown that the optimal contract has this simple form, we can treat the model as a static one and focus on time $t = 2$ realizations: $\tilde{R}_{n,\ell} = \tilde{R}_{n,\ell}(2)$, $\tilde{x}_n \stackrel{\text{def}}{=} \tilde{x}_n(2) = \tilde{x}_n(2)$, $\tilde{\varepsilon}_{n,\ell} \stackrel{\text{def}}{=} \tilde{\varepsilon}_{n,\ell}(2)$, and $e_n \stackrel{\text{def}}{=} \int_1^2 e_n(t)dt$. It follows that the time $t = 2$ revenue of firm ℓ in industry n is

$$\tilde{R}_{n,\ell} = \alpha_n I_{n,\ell} e_n + \tilde{x}_n I_{n,\ell} + \tilde{\varepsilon}_{n,l},$$

where $\tilde{x}_n$, $n = 1, \ldots, N$, are jointly normally distributed, with $\tilde{x}_n \sim N(1, \sigma^2_{x,n})$. Economy wide, these multivariate real risk factors are the $N \times 1$ vector, $\tilde{\boldsymbol{x}} = (\tilde{x}_1, \ldots, \tilde{x}_n)^T$, where $\tilde{\boldsymbol{x}} \sim N(\mathbf{1}, \Sigma)$, $\mathbf{1}$ is an $N \times 1$ vector of ones, and the element on the ith row and jth column of the $N \times N$ matrix Σ is $\sigma_{i,j}$. In what follows, we assume that the $N \times N$ covariance matrix, Σ, is non-singular; throughout most of the paper, we will allow for arbitrary covariance matrices. The firm-specific idiosyncratic risk $\tilde{\varepsilon}_{n,\ell}$ is also normally distributed, $\tilde{\varepsilon}_{n,\ell} \sim N(0, \sigma^2_{\varepsilon,n})$. The properties of these idiosyncratic shocks are summarized by the vector $\boldsymbol{\sigma}_\varepsilon = (\sigma_{\varepsilon,1}, \ldots, \sigma_{\varepsilon,N})^T$.

We place two restrictions on each firm's production possibilities. Suppose that firm ℓ in industry n raises and invests $I_{n,\ell}$. First, we assume that there is a convex investment cost. Specifically, in order to produce, the firm makes a sunk investment of $\kappa + \gamma\alpha_n I^2_{n,\ell}$, where $\kappa > 0$ and $\gamma > 0$ are constants.[3] Note that the second derivative of the cost function is $2\alpha\gamma$. Thus, for the

2. Two firms are in the same industry if they have the same production technology, not necessarily because they produce the same output.

3. The model is easy to generalize to industry-specific κ's and γ's. It is also straightforward to generalize the model to arbitrary means, $\tilde{\boldsymbol{x}} \sim N(\bar{\boldsymbol{x}}, \Sigma)$.

firms in which labour is more productive, the cost of installing capital is marginally more expensive. Overall, the cost function captures the fact that physical capital and financial capital are not perfectly fungible. We interpret this sunk cost as payments for research and development. The functional form is motivated by two stylized facts: marginal investment costs are increasing in investment level and research and development (R & D) costs per unit of investment are empirically higher in high-labour-productivity industries than in low-productivity ones.[4] This assumption, as we elaborate on later, is important in establishing a "size effect" and thereby for the asset pricing results in Section 4.7. For the general equilibrium results that we will derive, however, no specific functional relationship between α, κ, and γ needs to be assumed. Second, we assume that in order to produce, the worker–capital ratio is constant and equal to one. The restriction that one unit of investment requires one unit of workers makes the model tractable (in general equilibrium, two market-clearing conditions collapse into one). We note, however, that even though the worker–capital ratio is fixed, the effort–capital ratio is endogenous and depends on the specific industry. Implicitly, we assume that the workers of a firm act as one "representative worker", disregarding the intra-firm coordination problems that are surely present in large firms.

Firms raise investment capital in financial markets, which are open at time $t = 1$ and time $t = 2$ (these are the only times that firms have a strict incentive to participate in the markets). Firms raise capital by selling claims to their $t = 2$ profits. We assume that workers cannot trade in the stock of their own company, however they can trade in all other companies. Therefore, they can lay off the systematic risk that they are exposed to through their labour contract but not their idiosyncratic employment risk. Given this restriction, as all firms are identical and as systematic risk within each industry is perfectly correlated, without loss of generality, we can assume that there are N representative stocks each of which is perfectly correlated with industry risk. Prices for these stocks are determined by the interaction of workers who hedge consumption risk and the firms who raise capital. Stock n has price $S_n(t)$ at $t = 1, 2$. The return of stock n is denoted $\tilde{\mu}_n$, *i.e.* $S_n(2) = (1 + \tilde{\mu}_n)S_n(1)$. The random market returns can then be summarized by $\tilde{\boldsymbol{\mu}} \sim N(\bar{\boldsymbol{\mu}}, \Sigma_\mu)$, where $\bar{\boldsymbol{\mu}}$ is an $N \times 1$ vector of returns with typical element $\bar{\mu}_n$ and Σ_μ is the $N \times N$ covariance matrix. It will become clear that the equilibrium distribution of returns is normal and that in equilibrium, invertibility of Σ is equivalent to invertibility of Σ_μ. We therefore proceed under the assumption that Σ_μ is invertible. We define $\boldsymbol{\sigma}_{\mu,n} = \mathrm{cov}(\tilde{\boldsymbol{\mu}}, \tilde{x}_n)$. There is also a risk-free asset in perfectly elastic supply, with excess return normalized to zero.

Each investor working in industry n chooses a portfolio of dollar amounts in each industry denoted by $\boldsymbol{q}_n$, which is an $N \times 1$ vector describing his investment in each industry. At $t = 2$, his portfolio has value $\tilde{\theta}_n = \tilde{\boldsymbol{\mu}}^T \boldsymbol{q}_n$. For later convenience, we introduce the matrix $\boldsymbol{Q} = [\boldsymbol{q}_1, \ldots, \boldsymbol{q}_N]$.

In what follows, we focus on symmetric outcomes, so that all firms within the same industry offer the same contract to their employees and all employees in a particular industry invest in the same way. The exogenous parameters are given by the tuple $\mathcal{E} = (M, \kappa, \gamma, \Sigma, k, \rho, \boldsymbol{\alpha}, \boldsymbol{\sigma}_\varepsilon)$. Here, economy-wide human capital productivity is characterized by the vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)^T$. An equilibrium is characterized by the tuple of endogenous quantities: $\mathcal{X} = (w_0, \boldsymbol{L}, \boldsymbol{I}, \boldsymbol{e}, \boldsymbol{w}, \bar{\boldsymbol{\mu}}, \boldsymbol{Q})$, which we will define below. The first four elements of $\mathcal{X}$ constitute the real part of the economy.

The variable w_0 in an economy is the *ex ante* value of a worker's human capital. Specifically, it is the difference between the certainty equivalent of earned labour income and the disutility of working. As we demonstrate below, in the optimal contract, the worker retains this surplus and therefore w_0 can be thought of as the wage a worker would earn at a hypothetical firm

4. Growth firms—which are usually in high-productivity industries—tend to be R & D intensive even when adjusted for size, see, *e.g.* Chan, Lakonishok and Sougiannis (2001), hence the α_n-term in $\gamma \alpha_n I_{N,l}^2$.

that requires no effort from its workers. We will also refer to w_0 as the *effort-free wage* or the *participation wage*.[5]

The vector $\boldsymbol{L} = (L_1, \ldots, L_N)^T$ summarizes the distribution of firms across all sectors. Since there is a continuum of firms, each element is a positive real number, representing the total mass of firms in a specific industry. The initial size of each firm or alternatively investment in physical capital is given by $\boldsymbol{I} = (I_1, \ldots, I_N)^T$ and the workers' supply of human capital by $\boldsymbol{e} = (e_1, \ldots, e_N)^T$.

The incentive wages are given by $\boldsymbol{w} = (w_1, \ldots, w_n)^T$. A firm may not condition its payments on other firms' revenues (we assume that the firm observes its own performance before the performance of other firms becomes public, *i.e.* that there is delayed information dissemination) but has full freedom in designing a contract conditioned on its own revenue history, $\tilde{w}_{n,\ell}(t) = F(\{\tilde{R}_{n,\ell}(\tau)\} \mid 1 \leq \tau \leq t)$.[6]

Since all firms in an industry choose the same type of contract, there is a representative contract, $\tilde{w}_n$, in each industry.

For a given technology and exogenous structure of risk, the amount of risk generated will depend on these production choices and is therefore endogenous. The last two elements in $\mathcal{X}$ are the equilibrium financial variables: the expected return of each asset ($\bar{\boldsymbol{\mu}}$) and agents' portfolio choices ($\boldsymbol{Q}$).

Definition 1. General equilibrium of the economy $\mathcal{E}$ is characterized by $\mathcal{X}$ in which

(i) each firm optimally chooses an investment level and a wage contract to maximize expected profits leading to $\boldsymbol{I}$ and $\boldsymbol{w}$,
(ii) given a wage contract, each worker optimally chooses his effort level and stock market investment to maximize expected utility, leading to $\boldsymbol{e}$ and $\boldsymbol{q}$,
(iii) asset markets clear: $M\boldsymbol{q} = \text{diag}(\boldsymbol{I})\boldsymbol{L}$,
(iv) labour markets clear: $M = \boldsymbol{I}^T\boldsymbol{L}$,
(v) for each sector $n = 1, \ldots, N$, each firm makes zero economic profits,
(vi) $I_n > 0$, $L_n > 0$, $e_n > 0$ for all $n = 1, \ldots, N$.

Conditions (i) and (ii) are standard conditions that ensure that all economic agents are optimizing. Conditions (iii) and (iv) are market-clearing conditions; however, the second reflects our assumption that each worker is paired with a unit of capital, irrespective of the amount of effort he exerts. Thus, labour market clearing can be ensured with a condition on investment because the two factors are used in fixed proportion. An implication of condition (v) will be that the expected return in each sector of the financial market equals the cost of capital of that sector in the real economy, which will be an important step in "closing" the model. In this two-period world, a return higher than the cost of capital is equivalent to positive profits. Such rents are incompatible with general equilibrium as firms would enter into industries in which there are positive profits, driving rents to the fixed entry costs (κ) and thus generating zero economic

5. In constructing equilibrium, workers are indifferent between remaining in their sector and going elsewhere including this effort-free alternative.

6. If the information dissemination into the stock market is not immediate, it is natural for the firm, which pays wages continuously to its workers, to be better informed about its own value creation than about its competitors'. In the extreme case, when the stock market is only open at $t = 1$ and at $t = 2$, there is complete information delay. In the case when information diffusion is instantaneous, however, the firm may choose a compensation contract that filters out the systematic revenue component. In such a situation, the general equilibrium effects that we will derive still hold, but the asset pricing implications (discussed in Section 4.7) change. Specifically, the CAPM also holds in the moral hazard economy under the alternative specification.

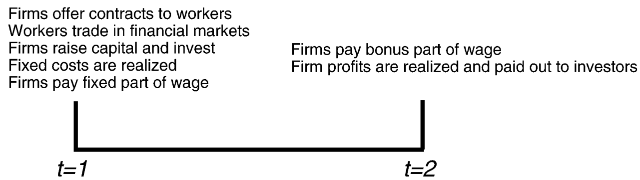

FIGURE 1
Sequence of events

profits. Finally, through condition (vi) we focus on outcomes that are interior: those for which investment, number of firms, and effort are all positive.

Central to our asset pricing results is the idea that systematic risk can be transferred either through the claims on the firms that the workers own (dividends) or through wages. Therefore, characterizing the optimal contract is important. To do so, we observe that our framework is similar (but not identical) to that presented in Holmström and Milgrom (1987), who demonstrate that the solution to the continuous contracting problem involves a fixed wage and a variable component that is linear in the realization of the firm's time $t = 2$ revenue. The major points of difference are that our firms observe revenues, which contain a systematic risk component, that the systematic component is priced in the market (and therefore the firm is not risk neutral with respect to this risk), and that workers with signed employment contracts can lay off any systematic risk in the stock market. Relative to Holmström and Milgrom (1987), each worker effectively has an endowment of systematic risk—his hedging portfolio. Of course, firms rationally anticipate this when they offer the employment contracts, and because they offer the contracts, the firms extract all rents above the workers' participation constraint. Also, since the worker can (*ex ante*) trade in the systematic risk, he will agree with the firm about its value and effectively hedge the risk in the market. We may therefore expect the argument of Holmström and Milgrom (1987) to go through within our setting.

Indeed, we show in the proof of the equilibrium existence, Proposition 1, that there is an equilibrium in which firms optimally choose a wage contract that contains a fixed wage and a variable component that is linear in the realization of the firm's time $t = 2$ revenue. For the time being, however, we *assume* that an industry-n contract yields wages of the form

$$\tilde{w}_{n,\ell} = s_n + b_n \times \tilde{R}_{n,\ell}. \tag{3}$$

That is, each firm pays a fixed wage, s_n, which we assume is front loaded (*i.e.* paid at $t = 1$ since it does not depend on any future information), and a wage that is linear in revenue with slope b_n, which necessarily needs to be back loaded (*i.e.* paid at $t = 2$). Given such a contract, it is immediately clear from the convexity of the cost of effort in the agent's utility function (1) that each worker will exert a fixed effort level at each instant in time, $e(t) = e$. We now proceed with the analysis under the linear contract assumption, which will later be shown to be optimal. The sequence of events are then as described in Figure 1.

3. PARTIAL EQUILIBRIUM ANALYSIS

We proceed in two steps. First, we suppose that a worker has been offered a contract and characterize his portfolio. This will allow us to determine his reservation utility with this optimal hedging. Given this, we then exhibit the optimal linear contract chosen by each profit-maximizing firm.

3.1. *Agents' optimal portfolios and effort levels*

Because the contract includes systematic risk, the risk-averse worker optimally chooses a portfolio that is negatively correlated with the firm's output. The agent chooses a portfolio $\tilde{\theta}_n$ (represented by the dollar portfolio $\boldsymbol{q}_n$) and an effort level e_n to maximize the certainty equivalent of the increase in her utility:

$$\Delta W_n = \max_{\boldsymbol{q}_n, e_n} \left[E(\tilde{\omega}_n + \tilde{\theta}_n) - \frac{k}{2} e_n^2 - \frac{\rho}{2} \mathrm{var}(\tilde{\omega}_n + \tilde{\theta}_n) \right]. \tag{4}$$

Lemma 1. *If firm n offers a linear contract* (s_n, b_n), *then*

(i) *the worker supplies effort* $e_n = \frac{b_n \alpha_n}{k}$,
(ii) *the worker holds a portfolio position*

$$\boldsymbol{q}_n = \Sigma_\mu^{-1} \left(\frac{\bar{\boldsymbol{\mu}}}{\rho} - b_n \boldsymbol{\sigma}_{\mu,n} \right). \tag{5}$$

Workers hedge the risk that they are exposed to through the incentive contract: *i.e.* they eliminate exposure to $\tilde{x}$ risk by shorting claims correlated with the industry in which they work (part (ii)). This mirrors the argument made in Bodie, Merton and Samuelson (1992) that even though human capital is not tradable, human capital risk may be partly hedgeable in the market. If $b_n = 0$, so that the firm does not offer an incentive component, all workers hold the same portfolio, $\boldsymbol{q}_n = \Sigma_\mu^{-1} \frac{\bar{\boldsymbol{\mu}}}{\rho}$. Thus, the term $b_n \Sigma_\mu^{-1} \boldsymbol{\sigma}_{\mu,n}$ in part (ii) of Lemma 1 is the distortion in portfolio holdings that comes about because the firm transfers wealth to the agent through labour income.[7]

In any equilibrium, all assets are long in the market portfolio (else markets do not clear). Therefore, going "short" is relative to the market portfolio and would be observed in the data as a distortion from the market portfolio and not necessarily as actual short positions. Aggregate data on agents' specific portfolio holdings are rare, but recent detailed data from Sweden are consistent with our model. Specifically, Campbell, Calvet and Sodini (2007) find that the median Swedish investor holds a well-diversified portfolio (as measured by a Sharpe ratio) but that there is considerable cross-sectional dispersion in investment efficiency which could be due to labour income. Finally, they find that sophisticated investors (measured as those with higher disposable income) are more likely to be underdiversified. This is consistent with our view that investors with larger incentive payments are more likely to distort their holdings from the market portfolio. We note, however, that a full test of the model would fully condition on wage income and incentive or bonus payments in addition to financial assets. To date, we know of no study that has done so.

3.2. *The firm's cost-minimizing contract*

In general equilibrium, if agents hold well-diversified portfolios, then the firm's shareholders will comprise all the agents in the economy (except the mass that are employed at the specific firm). Further, because all agents can trade in financial markets, the objective of each firm is well defined: in general equilibrium, all agents will agree on the opportunity cost of investment

7. That investments in such hedging portfolios arise when labour income risk is present was observed in Mayers (1973) and in several subsequent papers.

in that firm, and the cost of capital will reflect the systematic risk to which each shareholder is exposed.

To solve the firm's partial equilibrium problem, we let r_n be the risk-adjusted cost of capital in industry n. Given the cost of capital, the economic profits of a firm are

$$\tilde{\pi}_{n,\ell} = \underbrace{\alpha_n I_{n,\ell} \int_1^2 e_{n,\ell}(t)dt + \tilde{x}_n I_{n,\ell} + I_{n,\ell}\tilde{\varepsilon}_{n,\ell}}_{\text{Revenues}} - \underbrace{I_{n,\ell}\tilde{w}_{n,\ell}}_{\text{Cost of wages}}$$
$$- \underbrace{(\kappa + \gamma\alpha_n I_{n,l}^2)}_{\text{Production costs}} - \underbrace{r_n I_{n,\ell}}_{\text{Cost of capital}} . \tag{6}$$

Shareholders unanimously agree that the firm should maximize these profits.

A firm offering an optimal contract ensures that each worker's participation constraint is binding, *i.e.* that it appropriates any surplus above the participation constraint for all workers. This constraint reflects both the hedging that the workers will do in financial markets and the fact that they have outside employment options: agents will not work for a firm unless they are at least as well off as they would be if they worked in another industry or did not work at all. Therefore, they must be recompensed for their effort. The former will depend on the realization of the firm's revenues, while the latter will be the fixed wage. Note that the firms take into account the utility workers get from optimally trading in financial markets, by offering contracts that drive workers down to their participation constraint, *including* their portfolio holdings.

Lemma 2. *A firm in sector n offering an incentive wage b_n will offer the following fixed wage*

$$s_n(b_n, I_n) = \underbrace{\Delta W}_{\textit{outside option}} - \underbrace{\left(\frac{A}{2\rho} + b_n^2\left(\frac{\rho}{2}(C_n - \sigma_{x,n}^2)\right) - b_n(B_n - 1)\right)}_{\textit{surplus from market trading}}$$
$$+ \underbrace{b_n^2 \frac{\rho\sigma_{\varepsilon,n}^2}{2}}_{\textit{compensation for idiosyncratic risk}} - \underbrace{\frac{b_n^2\alpha_n^2}{2k}}_{\textit{surplus from exerting effort}}$$

where

$$A = \bar{\boldsymbol{\mu}}^T \Sigma_\mu^{-1} \bar{\boldsymbol{\mu}}, \quad B_n = \bar{\boldsymbol{\mu}}^T \Sigma_\mu^{-1} \boldsymbol{\sigma}_{\mu,n}, \quad C_n = \boldsymbol{\sigma}'_{\mu,n} \Sigma_\mu^{-1} \boldsymbol{\sigma}_{\mu,n}.$$

The term ΔW is the value of the worker's outside option, which is an important general equilibrium quantity. However, in partial equilibrium the firm and worker take it as given. To interpret the second term, observe that the constants A, B_n, and C_n reflect trade in the stock market. As the risk-free rate in this economy is normalized to zero, A represents the squared market Sharpe ratio. Thus, the term $\frac{A}{2\rho}$ is the certainty equivalent of market participation if the firms do not pay out wages. However, firms do pay out wages and the other terms reflect level and risk adjustments of trading off the systematic risk in the employment contract.

We note that, because of the normality of the risk distributions, realized compensation may be negative. This is unpalatable, however it is consistent with a more complicated labour market in which workers need to commit resources, such as buying equipment or getting training before joining a firm. Overall, this is a drawback of the CARA/normal framework.

The adjustment for idiosyncratic risk: $\frac{b_n^2 \rho \sigma_{\varepsilon,n}^2}{2}$ is direct compensation for the risk that the worker cannot lay off through trade in the stock market. This element of the moral hazard problem will allow us to derive a value effect in general equilibrium.

The final term $\frac{b_n^2 \alpha_n^2}{2k}$ reduces the fixed wage because if workers supply an effort level of $e_n = \frac{b_n \alpha_n}{k}$, they experience a disutility of $\frac{b_n^2 \alpha_n^2}{2k}$. However, in expectation the incentive part of the contract yields twice that: $b_n \alpha_n \frac{\alpha_n b_n}{k}$. The net surplus that the worker achieves of $\frac{b_n^2 \alpha_n^2}{2k}$ directly reduces the fixed wage so that he is driven down to his participation constraint.

We note, in passing, that since the risk-free asset is in elastic supply, the firm is not constrained at $t = 1$ but can choose arbitrary investment and wage levels. If capital is borrowed, however, it will affect the pay-offs at $t = 2$. When we discuss empirical results and in our predictions, we distinguish between wages (s_n), incentives (b_n), and total compensation w_n.

Given that the contract offered in industry n is characterized by the pair (b_n, s_n), which induces an effort level of e_n, the firm's profit function is

$$\tilde{\pi}_n = \underbrace{I_n(\alpha_n e_n + \tilde{x}_n)}_{\text{Revenue}}(1 - b_n) - I_n s_n - (\kappa + \gamma\, \alpha_n I_n^2) - r_n I_n. \tag{7}$$

Because a portion of the firm's revenue is paid out to workers, the wage bill acts as "leverage" on the revenue of the firm. In other words, part of the risks in the real economy are transferred to the workers through their incentive contracts. This is captured by the $(1 - b_n)$-term.

The same sort of "leverage" reasoning applies to a firm's cost of capital. As the shareholders who supply the capital only have claims on the portion of revenues not paid out to the employees in compensation (*i.e.* $\tilde{R}_n(1 - b_n)$), they view the opportunity cost of each dollar invested as something that generates claims to sector n risk that is reduced by a factor of $1 - b_n$. Thus, the cost of capital is also reduced by the risk that will be paid out to the workers. Specifically, suppose that in equilibrium, one unit of $\tilde{\boldsymbol{x}}_n$ risk commands a required rate of return, z_n. Further, a unit investment in industry n generates a unit of $\tilde{x}$ risk but only $(1 - b_n)$ accrues to the shareholders. In this case, the cost of capital is

$$r_n = (1 - b_n) z_n.$$

This follows from a no-arbitrage condition since the cost of capital for a risk-free investment is 0 and the payout to shareholders is a combination of risk-free capital and $\tilde{x}$ risk. If firms pay out all their risky cash flows to workers, the firm is risk free and the cost of capital falls to zero, the risk-free rate. If the firm pays out no incentive bonuses, then it retains the industry $\tilde{\boldsymbol{x}}$ risk and the cost of capital is maximal.

As individual firms take z_n as given, they choose physical investment and labour investment to maximize risk-adjusted expected profits:

$$\begin{aligned}\max_{I_n, b_n} E[\tilde{\pi}_n] &= \max_{I_n, b_n} I_n((\alpha_n e_n + 1)(1 - b_n) - s_n(b_n, I_n) - (1 - b_n) z_n - \alpha_n \gamma\, I_n) - \kappa \\ &= \max_{I_n, b_n} I_n \left(\left(\frac{b_n \alpha_n^2}{k} + 1 \right) (1 - b_n) - s_n(b_n, I_n) - (1 - b_n) z_n - \alpha_n \gamma\, I_n \right) - \kappa. \end{aligned} \tag{8}$$

The first-order condition for physical capital:

$$\frac{\partial E[\tilde{\pi}_n]}{\partial I_n} = \underbrace{\left(\frac{b_n \alpha_n^2}{k} + 1\right) - 2\alpha_n \gamma I_n}_{\text{value of the marginal product of capital}}$$

$$-\underbrace{\left(\left(\frac{b_n \alpha_n^2}{k} + 1\right)(b_n) + I_n \frac{\partial s_n(b_n, I_n)}{\partial I_n} + s_n(b_n, I_n) + (1 - b_n) z_n\right)}_{\text{change in cost}} = 0,$$

suggests that the firm equates the value of the marginal revenue product of capital (the first term), with the change in costs it incurs if it increases capital (the second term). The change in cost consists of three terms. The first is the effect on the wage bill of increasing investment. If the firm increases investment, then, *ceteris paribus*, the firm has to recompense existing workers for the fact that the value of the workers' shares is now lower. Second, an increase in investment induces an increase in the labour force and therefore the firm pays out an extra fixed wage. Finally, for every increase in investment, the firm pays the direct cost of $\tilde{x}$ risk in the capital markets.

Solving the two first-order conditions yields the firm's optimal decision.

Lemma 3. *Firms in sector n will choose*

$$b_n = \frac{\alpha_n^2/k + z_n - b_n}{\alpha_n^2/k + \rho(\sigma_{x,n}^2 - C_n + \sigma_{\varepsilon,n}^2)},$$

$$I_n = \frac{1}{4\alpha_n \gamma_n}\left(\frac{(\alpha_n^2/k + z_n - b_n)^2}{2[\alpha_n^2/k + \rho(\sigma_{x,n}^2 - C_n + \sigma_{\varepsilon,n}^2)]} + \frac{A}{2\rho} - \Delta W + 1 - z_n\right),$$

if $I_n > 0$ *and* $0 < b_n < 1$.

Recall that workers supply effort increasing in b_n but by a factor of $\frac{\alpha_n}{k}$. Thus, Lemma 3 can also be viewed as a firm's choice of human capital and physical capital. We define the vectors $\boldsymbol{b} = (b_1, \ldots, b_n)$ and $\boldsymbol{s} = (s_1, \ldots, s_n)$.

The partial equilibrium choices of the firm do not give any insight into the cross-sectional attributes of either production or employment contracts. Consider the incentive term b_n that determines the proportion of revenues that are paid out to the worker. From Lemma 3, the larger the moral hazard problem (measured by $\sigma_{\varepsilon,n}$) the less steep the incentive schedule. However, it is unclear how this incentive payment varies with different firm productivities: more productive firms may offer smaller incentive payments if the cost of providing such incentives is sufficiently high.

The relationship between the size of each firm and labour productivity is also not clear from inspection of Lemma 3. Tedious derivation yields that firm size can either be increasing or decreasing in productivity. Intuitively, if the fixed size of investment is sufficiently large, then increasing productivity can increase or decrease the optimal size of the firm.

4. GENERAL EQUILIBRIUM

In general equilibrium, all markets clear and so the prices of all real and financial assets are determined by supply and demand. The demand for the assets comes from the return opportunities they offer, as well as from the hedging motives of the agents, and the supply of assets comes from firms' optimal investment decisions. Therefore, there is an equilibrium link between agents' valuation for all financial assets and the underlying firms' cost of capital. Although these links are arbitrary in partial equilibrium, they can be completely characterized in general equilibrium and are pinned down if all firms earn zero economic profits.

Lemma 4. *If firms make zero expected profits, then in equilibrium*

(i) *the expected return on financial assets equals the opportunity cost of capital in the real economy, so that* $\bar{\mu}_n = r_n$,

(ii) *both the expected returns and the covariance of returns reflect real risks so that*

$$[\Sigma_\mu]_{i,j} = (1-b_i)[\Sigma]_{i,j}(1-b_j) \quad \textit{and} \quad (\boldsymbol{\sigma}_{\mu,i})_j = (1-b_j)[\Sigma]_{i,j}. \tag{9}$$

Another way of stating Lemma 4 is that in general equilibrium, if firms earn zero profits, then all real assets are priced based on the risk to which they are exposed. As we show later when we discuss our asset pricing implications, if the firm pays out part of its systematic risk in wages, then an econometrician who only observes the stock market may draw incorrect inferences about a firm's risk. Of course, no agents in the economy misprices risk and so how such cash flows are paid out is irrelevant. Hence, for a given level of moral hazard, how systematic risk is paid out (*i.e.* through the wage channel or not) does not affect the real part of the economy as all agents agree on the opportunity cost of investment in each industry.

We observe that the multivariate normality of $\tilde{\boldsymbol{x}}$ implies multivariate normality of $\tilde{\boldsymbol{\mu}}$, and as long as $b_n < 1$, $n = 1, \ldots, N$, invertibility of Σ is equivalent to invertibility of Σ_μ, as mentioned in Section 2.

The model's structure is shown in Figure 2. The exogenous parameters arise from industrial production characteristics, part A in the top of the figure. Our goal is to study how these primitive characteristics affect the equilibrium outcome of the other variables in the economy (parts B–D) and how, in equilibrium, the other variables are related.

An important variable from the real economy is the marginal total productivity of capital evaluated at the equilibrium level of investment:

$$\upsilon_n = 1 + b_n\alpha_n^2/2k - 2\sqrt{\kappa\gamma\,\alpha_n}, \quad n = 1, \ldots, N. \tag{10}$$

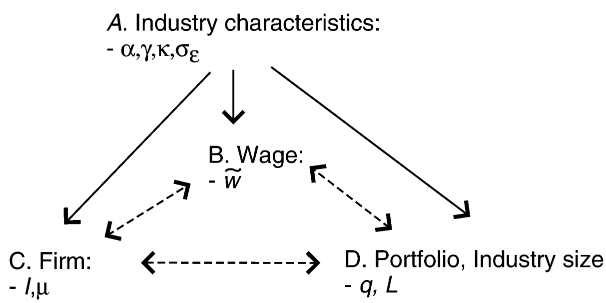

FIGURE 2

Model's relationships. Industry characteristics (productivity, α; entry costs, γ and κ; and degree of moral hazard, σ_ε) in equilibrium determine firm characteristics (size, I, and expected returns, μ), labour markets (wage contracts, $\tilde{w}$), and investments (portfolios, $\boldsymbol{q}$, and industry sizes, L).

When we formally develop our asset pricing results, we describe exactly how v_n is related to the firm's first-order condition for investment capital evaluated at the equilibrium investment level and effort-free wage. We define the vector $\boldsymbol{v} = (v_1, \ldots, v_n)^T$.

Finally, we impose various parameter restrictions to ensure the existence of an interior equilibrium.

Assumption 1.

(i) *The risk aversion of investors is sufficiently low. Specifically,*

$$\rho < \mathbf{1}^T \Sigma^{-1} \boldsymbol{v}.$$

(ii) *The risk aversion of investors is sufficiently high. Specifically,*

$$\rho \Sigma^{-1} \mathbf{1} > (\mathbf{1}^T \Sigma^{-1} \boldsymbol{v}) \Sigma^{-1} \mathbf{1} - (\mathbf{1}^T \Sigma^{-1} \mathbf{1}) \Sigma^{-1} \boldsymbol{v}. \tag{11}$$

The first part of Assumption 1 ensures that all workers are employed in equilibrium. If this condition does not hold, then investors are too risk averse to want to absorb the risk inherent in a full-employment equilibrium. Thus, they are better off at lower production and concomitant risk levels. The condition is always satisfied if there is at least one sector with very low risk. In this case, workers can always contribute to total surplus by working in a low-risk sector.

The second part of the assumption guarantees that agents are sufficiently risk averse so that they will not wish to hold negative positions in any of the assets in equilibrium. Equivalently, in general equilibrium this assumption ensures that investment in all firms is strictly positive. If this condition were violated (suppose that all agents were risk neutral), each would optimally invest arbitrary large amounts in the asset with the highest mean return and short all assets with lower returns. In this case, the markets for physical investment would not clear.

4.1. *The real economy under moral hazard and first best*

We are now in a position to prove the existence of a unique general equilibrium.[8] To make our analysis more accessible, we first describe the real economy with moral hazard and compare it to the first best and then turn to financial market equilibrium.

Proposition 1. *In an economy, $\mathcal{E}$, that satisfies Assumption 1, there is a unique equilibrium in which*

(i) *effort levels are $e_n = b_n \alpha_n / k$,*
(ii) *investment is $I_n = \sqrt{\frac{\kappa}{\gamma \alpha_n}}$,*
(iii) *the optimal incentive part of the wage contract is $b_n = \frac{1}{1 + \frac{k \rho \sigma^2_{\varepsilon,n}}{\alpha^2_n}}$,*
(iv) *the fixed part of the wage contract is $s_n = (1 - b_n) w_0 + \frac{b^2_n \rho \sigma^2_{\varepsilon,n}}{2} - 2 b_n \sqrt{\kappa \gamma \alpha_n}$,*
(v) *the effort-free wage is $w_0 = \frac{\mathbf{1}^T \Sigma^{-1} \boldsymbol{v} - \rho}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}}$,*

8. Strictly speaking, the equilibrium is unique among equilibria in which all firms optimally choose linear compensation contracts. In the proof, we show that it will indeed be optimal for a firm to offer a linear compensation contract, given that (almost) all other firms offer linear compensation contracts. We cannot rule out other equilibria in which many firms offer non-linear compensation contracts.

(vi) *the number of firms in each sector is* $\boldsymbol{L} = M\rho^{-1}\Lambda_I^{-1}\Sigma^{-1}(\boldsymbol{v} - w_0\mathbf{1})$,
(vii) *agents' certainty equivalent is* $W + w_0 + W_C$, *where* $W_C = (\boldsymbol{v} - w_0\mathbf{1})^T\Sigma^{-1}(\boldsymbol{v} - w_0\mathbf{1})/2\rho$.

In Proposition 2, we provide a formula for $\bar{\mu}_n$ (and thereby, via equations (5) and (9), for $\boldsymbol{Q}$), so the equilibrium outcome, $\mathcal{X}$, is completely specified. We note that (vii) provides a natural decomposition of the welfare gains to agents from labour markets (w_0) and capital markets (W_C).

The only friction in this economy is the idiosyncratic risk that the worker bears because he cannot short sell his own firm. He can, of course, short sell his own industry and so exposure to systematic risk does not affect welfare or economic efficiency. Thus, when $\sigma^2_{\varepsilon,n} = 0$, the economy achieves first best.[9] Intuitively, if the unobservable idiosyncratic risk is removed, if all agents optimize and there are no frictions, then the competitive equilibrium achieves first best.

Corollary 1. *In an economy,* $\mathcal{E}$, *that satisfies Assumption 1, in which there is no moral hazard* ($\sigma_{\varepsilon,n} = 0$)*, there is a unique equilibrium in the real economy in which*

(i) *effort levels are* $e_n = \alpha_n/k$,
(ii) *investment is* $I_n = \sqrt{\frac{\kappa}{\gamma\alpha_n}}$,
(iii) *the effort-free wage is* $\hat{w}_0 = \frac{\mathbf{1}^T\Sigma^{-1}\hat{\boldsymbol{v}} - \rho}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}}$,
(iv) *the number of firms in each sector is* $\hat{\boldsymbol{L}} = M\rho^{-1}\Lambda_I^{-1}\Sigma^{-1}(\hat{\boldsymbol{v}} - \hat{w}_0\mathbf{1})$,
(v) *agents' certainty equivalent is* $W + \hat{w}_0 + \hat{W}_C$, *where* $\hat{W}_C = (\hat{\boldsymbol{v}} - \hat{w}_0\mathbf{1})^T\Sigma^{-1}(\hat{\boldsymbol{v}} - \hat{w}_0\mathbf{1})/2\rho$.

Here,

$$\hat{v}_n = 1 + \alpha_n^2/2k - 2\sqrt{\kappa\gamma\alpha_n}, \quad n = 1, \ldots, N,$$

is the equilibrium marginal total productivity of capital in sector n of the friction-free economy and $\hat{\boldsymbol{v}} = (\hat{v}_1, \ldots, \hat{v}_n)^T$. Comparing the two definitions, the difference between $\mathbf{v}$ (moral hazard case) and $\hat{\mathbf{v}}$ (first best) is thus that in the friction-free economy, $b_n = 1$ in all sectors.

The most important difference between the first-best and moral hazard economy is that, relative to the first-best economy, equilibrium effort levels in the moral hazard economy are distorted by the incentive payment, b_n. This variable is driven by moral hazard in that the larger the moral hazard problem (as measured by σ_ε) the smaller the equilibrium value of b (Proposition 1 part (iii)). In the presence of moral hazard, it becomes expensive for a firm to provide incentives to a risk-averse agent and, therefore, firms in equilibrium provide fewer. This means that the equilibrium effort level elicited is lower (Proposition 1 part (i)), a direct social cost of the idiosyncratic risk borne by the agent.

Of particular interest to us is that in general equilibrium, the incentive payment b_n is increasing in a firm's productivity as evinced by Proposition 1 part (iii). In partial equilibrium, this sign was ambiguous. However, in general equilibrium, firms and sectors that are more productive will elicit higher effort levels from their workers. This means that they will pay out more risk through the wage channel and therefore will be under-represented in the market portfolio. That is, their size in the market portfolio is smaller than their actual size in the economy. Thus, the size of b_n is important for asset pricing because it affects the "risk leverage"—the additional risk agents are exposed to through labour markets—and thereby both the returns and the cost of capital.

Equilibrium firm investment levels, or equivalently firm size, are not distorted by moral hazard. This is evident from a comparison of Proposition 1 part (ii) with Corollary 1 part (ii).

9. In an extended version of this paper, we work through the details of the economy with no moral hazard. The analysis is available upon request.

However, even though the size of each firm in each industry is not affected by moral hazard, the overall size of each industry is, since the number of firms changes as seen by comparing Proposition 1 part (vi) with Corollary 1 part (iv). Intuitively, if all else is held equal, the extra production cost induced by moral hazard reduces the productivity of an industry. This affects agents' risk-return trade-offs and thereby equilibrium industry sizes. This also has asset pricing implications because differences in industry size across the two economies lead to a different equilibrium risk structure. We explore this in more detail when we present our asset pricing results.

4.2. *Moral hazard and financial markets*

It is straightforward to show that the CAPM holds when there is no moral hazard. This follows standard analysis in which the risks in the economy are generated by the real economy described in Corollary 1. For this given set of real risks, expected returns are described by the CAPM:

$$\bar{\mu}_n = \beta_n E[\tilde{\mu}_{\text{market}}], \quad \text{where } \beta_n = \frac{\text{cov}(\tilde{\mu}_n, \tilde{\mu}_{\text{market}})}{\text{var}(\tilde{\mu}_{\text{market}})} \text{ and } \tilde{\mu}_{\text{market}} = M^{-1}\sum_{n=1}^{N} I_n L_n \tilde{\mu}_n.$$

For the equilibrium set of real risks, this is a standard CAPM result: all assets are priced by their covariance with the market portfolio.

In the economy with moral hazard, the CAPM does not hold. This is primarily because the financial market alone no longer reflects the true underlying risks in the economy. Rational agents price all risks (irrespective of how they are paid out) and therefore returns seem "distorted" relative to a standard CAPM. Thus, beyond the fact that the real risks are different in the two economies, the pricing relationship between risk and asset returns also differs.

Proposition 2. *In an economy that satisfies the assumptions of Proposition 1, expected returns are* $\bar{\mu}_n = (1-b_n)(\upsilon_n - w_0)$. *Moreover, assume that the value-weighted market portfolio is* $\boldsymbol{q}$. *Define the diagonal matrix* $\Lambda = \text{diag}(\alpha_1/\sigma_{\varepsilon,1}, \ldots, \alpha_N/\sigma_{\varepsilon,N})$. *Then,*

$$\bar{\boldsymbol{\mu}} = \boldsymbol{\beta}\nu, \tag{12}$$

where

$$\boldsymbol{\beta} = \frac{\Sigma_\mu\left(\bar{\boldsymbol{I}} + \frac{1}{k\rho}\Lambda^2\right)\boldsymbol{q}}{\boldsymbol{q}^T\left(\bar{\boldsymbol{I}} + \frac{1}{k\rho}\Lambda^2\right)\Sigma_\mu\left(\bar{\boldsymbol{I}} + \frac{1}{k\rho}\Lambda^2\right)\boldsymbol{q}}, \quad \nu = \boldsymbol{q}^T\left(\bar{\mathbf{I}} + \frac{1}{k\rho}\Lambda^2\right)\bar{\boldsymbol{\mu}}, \tag{13}$$

and $\bar{\boldsymbol{I}}$ *is the* $N \times N$ *identity matrix.*

As this is a production economy, financial returns are explicitly related to real production variables. By inspection of the first-order condition for investment capital evaluated at the equilibrium effort level, the equilibrium value of the marginal product of investment capital is $\upsilon_n - w_0$. Increasing investment has a direct effect on profits measured by υ_n: however, in addition, the firm has to hire an extra worker which decreases profits by the minimum participation wage, w_0. Thus, stock market returns reflect the productivity of a marginal dollar invested in an industry. Note that υ_n is the value of the marginal product of capital plus the participation wage. In other words, it is the equilibrium *marginal total productivity*—the value generated to all stakeholders (shareholders and workers) of an extra unit of investment capital. Finally, the factor $1-b_n$ is just a firm-specific wage "risk-leverage" effect.

To recover a CAPM result, observe that equation (12) shows that expected excess returns are given by the product of a market return and a "beta factor". However, the market portfolio and betas are defined with respect to a modified portfolio

$$\tilde{\nu} = \frac{\boldsymbol{q}^T\left(\bar{\mathbf{I}} + \frac{1}{k\rho}\Lambda^2\right)\tilde{\boldsymbol{\mu}}}{\boldsymbol{q}^T\left(\bar{\boldsymbol{I}} + \frac{1}{k\rho}\Lambda^2\right)\mathbf{1}}. \tag{14}$$

This portfolio measures the true risk in the economy, taking into account the risk that is paid out through wages. The return on the market portfolio, on the other hand, is $\tilde{\mu}_{\text{market}} = M^{-1}\sum_{n=1}^{N} I_n L_n \tilde{\mu}_n$. In line with the argument in Roll (1977), we see deviations from the CAPM when we measure each industry's expected return with respect to $\bar{\mu}_{\text{market}}$. Proposition 2 can then be rewritten in the CAPM-like form

$$\bar{\mu}_n = \beta_n E[\tilde{\nu}], \quad \text{where } \beta_n = \frac{\text{cov}(\tilde{\mu}^n, \tilde{\nu})}{\text{var}(\tilde{\nu})}. \tag{15}$$

These deviations are typically non-linear in the underlying firm characteristics, which makes it difficult to obtain a simple closed-form expression. However, there is sufficient monotonicity to be able to construct empirical tests based on sorted portfolios. We elaborate on this in Section 4.7 below.

In what follows, we assume that both the moral hazard economy and the friction-free economy have interior equilibria, *i.e.* that Assumption 1 is satisfied, both for $\mathbf{v}$ and for $\hat{\mathbf{v}}$.

4.3. *Welfare from participation in both labour and asset markets*

It is immediate that the total welfare in the moral hazard economy is always lower than that in the friction-free economy; risk-averse agents are exposed to idiosyncratic risk in the former but not in the latter. However, the welfare that accrues to the agent from participating in either the labour or the financial market can be higher or lower in the economy with moral hazard. From Proposition 1 part (vii) and Corollary 1 part (v), we can decompose the welfare change (over endowments) into gains they receive from participating in the labour market (w_0) and the benefit they receive from participating in the capital markets (W_C).

From Proposition 1 and Corollary 1, one can construct a condition on the risk-weighted marginal productivity of capital that determines whether the welfare obtained through the labour market is higher or lower in the moral hazard economy.

Proposition 3. *The effort-free wage is higher in the moral hazard economy, $w_0 > \hat{w}_0$, if and only if $\mathbf{1}^T\Sigma^{-1}(\hat{\boldsymbol{v}} - \boldsymbol{v}) < 0$.*

The joint condition on productivity and sector risk ensures that investors respond to a change in real productivity by increasing their hedging demand, which leads to an increase in overall investment and drives up the participation wage.

To see the intuition for this condition, we compare an economy with moral hazard in only one sector to a completely friction-free economy. Specifically, consider an economy in which the noisiness is very low in all industries but one, $\sigma_{\varepsilon,n} > 0$, $\sigma_{\varepsilon,j} \approx 0$, $j \neq n$. In this economy, $[(\hat{\boldsymbol{v}} - \boldsymbol{v})]_n > 0$ and $[(\hat{\boldsymbol{v}} - \boldsymbol{v})]_j \approx 0$, for $j \neq n$. In other words, the marginal productivities of capital are roughly equal in all sectors except for one in which the marginal productivity of capital in the economy with moral hazard is strictly lower. Such a situation could occur if an economy without moral hazard was struck by a "noise shock" in sector n, which lead to a new equilibrium. (A shock could come about because of changes in industry-specific regulations that lead to uncertainty about worker input.)

In this comparison, given Proposition 3 and our assumptions about productivity, a necessary and sufficient condition for $w_0 - \hat{w}_0 > 0$ is

$$\sum_{j=1}^{N} [\Sigma^{-1}]_{n,j} < 0. \tag{16}$$

For example, the following covariance matrix satisfies equation (16), for the first sector, $n = 1$:

$$\Sigma = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & 0 \\ 2 & 0 & 3 \end{bmatrix}. \tag{17}$$

After the noise shock, real productivity is lower in the first sector. Or, holding everything else fixed, the real production in that sector falls. It immediately follows, holding everything fixed, that expected returns fall in that sector. However, given the hedging demand driven by condition (16), lower expected returns lead to higher total dollar demand in the stock market.[10] Intuitively, risk-averse investors compensate for the decrease in expected returns in sector 1 by investing more aggressively in the other two sectors.

Hence, in the "noisy" economy after a shock, demand for stocks is higher than supply. This leads firms to compete for labour, driving up wages. The wage increases further decrease the returns in the affected sector which has a multiplicative effect on stock demand. The wage increase also decreases expected returns in all other sectors which decreases the demand pressure on stocks and more than offsets the effects in sector i. A new equilibrium is reached in which w_0 is higher than before the noise shock and therefore higher than $\hat{w}_0$.

A somewhat different intuition underpins situations in which the welfare from capital market participation is higher in the moral hazard economy. These cases rest on how diversified the real risks in the economy are. Suppose equilibrium sector sizes are "unbalanced" in the real economy when there are no frictions. A moral hazard problem, that in equilibrium causes one of the industry sizes to shrink, may give a higher certainty equivalent to agents' stock market participation.

The condition is derived from Proposition 1 part (vii) and Corollary 1 part (v).

Proposition 4. *The welfare from capital market participation is higher in the moral hazard economy, $W_C > \hat{W}_C$, if and only if $(\boldsymbol{v} - \hat{\boldsymbol{v}} - (w_0 - \hat{w}_0)\mathbf{1})^T \Sigma^{-1} (\boldsymbol{v} + \hat{\boldsymbol{v}} - (w_0 + \hat{w}_0)\mathbf{1}) > 0$.*

A specific numerical example that has this effect can be built from the covariance matrix in equation (17). Further, suppose that $\hat{\boldsymbol{v}} = [3{\cdot}1, 2{\cdot}5, 2{\cdot}5]^T$ and that the risk-aversion parameter is $\rho = 1$. From Corollary 1 parts (v) and (vii), it is immediate that $\hat{W}_C = 0{\cdot}86$. If moral hazard is introduced in the second sector, so that $\boldsymbol{v} = [3{\cdot}1, 2{\cdot}4, 2{\cdot}5]^T$, then the new welfare from the capital market increases to $W_C = 0{\cdot}92$, *i.e.* $0{\cdot}06$ higher than in the friction-free equilibrium. The welfare loss in the labour market is $0{\cdot}1$ (it decreases from $0{\cdot}9$ to $0{\cdot}8$) so the total welfare change is $0{\cdot}06 - 0{\cdot}1 = -0{\cdot}04$. That is, the total welfare is lower in the moral hazard economy, as expected.

The intuition lies in the difference between the equilibrium sector sizes in the two economies. In the moral hazard economy, sectors are more balanced than in the friction-free economy. Under the first best, industry sizes are $\Lambda_I \boldsymbol{L} = [0{\cdot}2, 0{\cdot}4, 0{\cdot}4]^T$. Here, the first sector is dominated and

10. This is because the total dollar demand from each investor is $D = \mathbf{1}^T \boldsymbol{q} = \mathbf{1}^T \Sigma_\mu^{-1} \bar{\boldsymbol{\mu}}$. This increases if there is a negative shock in $\bar{\mu}^i$ since $\partial D / \partial \bar{\mu}^i = \sum_{j=1}^{N} [\Sigma_\mu^{-1}]_{i,j} < 0$.

so there is limited risk diversification. By contrast, in the moral hazard economy, the sector sizes are $\Lambda_I \hat{\boldsymbol{L}} = [0{\cdot}3, 0{\cdot}33, 0{\cdot}37]^T$ so the sector sizes are more balanced and the economy therefore offers more diversification benefits.

The overall inference that we can draw from these two examples is that although the total welfare change in the economy is unambiguous when moral hazard is introduced (it has to go down), the certainty equivalence of participation in any individual market can go up or down. Therefore, any partial equilibrium analysis of the effect of moral hazard in a specific market may be misleading.

4.4. *Moral hazard and industry size*

In partial equilibrium models of firm-financing frictions, moral hazard invariably reduces investment. It is *a priori* unclear whether such an intuition also holds in general equilibrium. As we have just argued, increased moral hazard in a sector changes the equilibrium effort-free wage. It also, through the wage contract, changes the portfolio investment decisions of workers. Yet, in spite of these general equilibrium effects, we establish that increased moral hazard in a sector always decreases the equilibrium size of that sector for any parameterization of the economy.

To establish this result, consider an economy in equilibrium, with total marginal productivity $\boldsymbol{v}$. Compare it to an otherwise identical economy, but with more moral hazard in one sector, n, *i.e.* with a higher $\sigma_{\varepsilon,n}$ all else equal. From Proposition 1 (iii), it follows that the incentive payment, b_n, is lower in the second economy, which in turn implies that the total marginal utility, υ_n, is lower. Indeed, it follows directly from equation (10) that the difference in υ_n between the two economies is $\Delta\upsilon_n = \frac{\alpha_n^2}{2k}\Delta b_n$. In other words, changes in b_n and υ_n are directly proportional.[11]

The total size of the sectors in the initial economy is $\Lambda_I \boldsymbol{L}$. It is straightforward to show (using Proposition 1 part (vi)) that the change in size can be written as

$$\Delta(\Lambda_I \boldsymbol{L}) = M\rho^{-1}\boldsymbol{X}\Delta\boldsymbol{v}, \quad \text{where } \boldsymbol{X} = \left(\bar{\boldsymbol{I}} - \frac{\Sigma^{-1}\mathbf{1}\mathbf{1}^T}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}}\right)\Sigma^{-1} \text{ is an } N \times N \text{ matrix.} \tag{18}$$

For each of the economies, the equilibrium is interior and therefore all resources are used and the size of the two economies is the same, *i.e.* $\mathbf{1}^T\Delta\boldsymbol{L} = 0$. It follows from equation (18) that the n-th sector decreases in size when moral hazard increases if and only if the nth diagonal element of $\boldsymbol{X}$ is positive, $[\boldsymbol{X}]_{n,n} > 0$. For any economy (of at least two sectors), this will always be true as Σ is a positive-definite matrix.

We have the following result.

Proposition 5. *The size of sector n in an economy in which sector-n moral hazard is higher is strictly smaller than the size in the original economy. Or,* $[\Delta(\Lambda_I \boldsymbol{L})]_{n,n} < 0$.

Therefore, the partial equilibrium intuition that higher moral hazard in a sector decreases its efficiency and so renders it smaller is robust to general equilibrium under general conditions. This automatically implies that the distribution of real risks in the economy with moral hazard differs from the first-best economy.

4.5. *Industry size and the cost of capital*

Changes in moral hazard affect industry size, but are changes in industry size related to the cost of capital? A standard asset pricing intuition is that if a sector is (*ceteris paribus*) larger,

11. In vector notation, the change can be written as $\Delta\boldsymbol{v}$, where all elements of $\Delta\boldsymbol{v}$ are zero, except for the n-th element, $\Delta\boldsymbol{v} = \frac{\alpha_n^2}{2k}\Delta b_n < 0$.

in equilibrium, its expected return must be higher to recompense risk-averse investors for the increased undiversifiable risk. Such an argument is the basis for the wealth puzzle documented in Bansal, Fang and Yaron (2006). However, this intuition may not hold under moral hazard.

To show how a counterexample naturally arises, we simplify the real risks in the economy and characterize an economy in which the real risks have a symmetric one-factor structure. If risks have a symmetric one-factor structure, then the real risks in the economy can be expressed as $\Sigma = c_0 \bar{\boldsymbol{I}} + c_1 \mathbf{1}\mathbf{1}^T$, $c_0 > 0$, $c_1 > 0$, and $\bar{\boldsymbol{I}}$ is the identity matrix, while $\mathbf{1}$ is a vector of ones. In such a world, covariances are positive and equal to c_1, while variances are simply $c_0 + c_1$. First, consider changes in industry size brought about by changes in productivity.

Lemma 5. *If risks in an economy have a one-factor structure, then a productivity shock in industry n affects own industry size by an amount*

$$\frac{d(L_n I_n)}{d\alpha_n} = \left(1 - \frac{1}{N}\right) \times \frac{M}{\rho c_0} \times \frac{\partial v_n}{\partial \alpha_n}$$

and other industries ($p \neq n$) by an amount

$$\frac{d(L^p I^p)}{d\alpha_n} = -\frac{1}{N} \times \frac{M}{\rho c_0} \times \frac{\partial v_n}{\partial \alpha_n}.$$

Changes in all industry sizes are driven by changes in firms' equilibrium total productivity $\left(\frac{\partial v_n}{\partial \alpha_n}\right)$. If the productivity shock increases the marginal product of capital, then the industry becomes larger and other industries shrink. The condition for a productivity shock to increase equilibrium total productivity is $\kappa\gamma < \frac{b_n^2 \alpha_n^3}{k^2}$. This condition demands that the industry cost of entry be low relative to productivity. If this holds, then changes in relative productivity change the size of each sector as capital flows to the more productive sectors in the economy.

These changes do not automatically lead to an increase in the cost of capital. The change in labour productivity has two effects on that sector's cost of capital. First, total productivity (v_n) increases, but second there is an effect on the effort-free wage, w_0. The second effect is a general equilibrium one, and through it, a change in the productivity of one sector will also affect the cost of capital in other sectors. If the *ex ante* value of human capital is higher, this affects firms of different productivity to different degrees, and therefore the distribution of risk in the economy. Of course, a risk-averse agent's valuation for any risk factor depends on others to which he is exposed.

To make this point more precise, consider the effect on the sector p's cost of capital of a change in the productivity of sector n.

Corollary 2. *If risks in the economy have a one-factor structure, then a productivity shock in sector n affects the cost of capital in sector n by an amount*

$$\frac{dr_n}{d\alpha_n} = (1 - b_n)\left(1 - \frac{1}{N}\right) \times \frac{\partial v_n}{\partial \alpha_n} - \frac{\partial b_n}{\partial \alpha_n}(v_n - w_0) \tag{19}$$

and the cost of capital in sector $p \neq n$ by an amount

$$\frac{dr_p}{d\alpha_n} = -\frac{1-b}{N} \times \frac{\partial v_n}{\partial \alpha_n} \quad p \neq n. \tag{20}$$

Note that the sign of $dr_p/d\alpha_n$ is always opposite to that of $\partial v_n/\partial\alpha_n$. By contrast, the effect of a productivity shock on the same sectors' cost of capital ($dr_n/d\alpha_n$) is ambiguous. Indeed, if $\frac{1-b_n}{\partial b_n/\partial\alpha_n} < \frac{1-1/N}{v_n - w_0}$, then $\partial r_n/\partial\alpha_n$ has the same sign as $\partial v_n/\partial\alpha_n$, otherwise it has the opposite sign.[12]

It is clear that situations can arise in which $\frac{\partial r_n}{\partial\alpha} > 0$ $\left(\text{because } \frac{\partial v_n}{\partial\alpha_n} < 0\right)$, in which case

$$\frac{d(L_n I_n)/d\alpha_n}{dr_n/d\alpha_n} < 0.$$

Thus, in the economy with moral hazard, the standard asset pricing intuition—that if one sector increases its size, its returns must increase—may not hold. The intuition for this is clear: a productivity increase leads to a larger sector, but more risk is then paid out through the wage channel, which *decreases* the cost of capital observed in the capital market.

4.6. *Welfare effects of productivity shocks*

Productivity shocks increase the size of a sector if the shock increases the equilibrium marginal product of capital in that sector. Under what conditions do productivity shocks increase aggregate welfare? In the absence of moral hazard, agents are better off as long as $\partial v/\partial\alpha > 0$. This is because, even though all workers are driven down to their endogenous participation constraint, employment is more productive; implicitly, this increases the *ex ante* value of human capital. However, in the presence of moral hazard, the conditions under which productivity shocks increase welfare are different. In particular, a shock that increases welfare in the friction-free economy may lead to a decrease in welfare if there is moral hazard present. Indeed, we have the following proposition.

Proposition 6. *A productivity shock is more likely to decrease welfare in the moral hazard economy than in the friction-free economy.*

It is easiest to see this by rewriting welfare in terms of the marginal productivity of capital. The increase in agents' welfare is the sum of the effort-free wage plus the value of investing in the stock market. Specifically, from the relationship that we established in Corollary 1 part (v) welfare in a friction-free economy is $\widehat{\Delta W} = \hat{w}_0 + (\hat{\mathbf{v}} - \hat{w}_0\mathbf{1})^T\Sigma^{-1}(\hat{\mathbf{v}} - \hat{w}_0\mathbf{1})/2\rho$. Therefore, the increase in welfare changes with productivity according to

$$\frac{d(\widehat{\Delta W})}{d\alpha^n} = \frac{d\hat{w}_0}{d\alpha^n} + \frac{1}{\rho}(\hat{\mathbf{v}} - \hat{w}_0\mathbf{1})^T\Sigma^{-1}\left(\frac{d\hat{\mathbf{v}}}{d\alpha^n} - \frac{d\hat{w}_0}{d\alpha^n}\mathbf{1}\right).$$

To simplify this further, observe that in equilibrium, the total size of a sector equals the total investment of the M investors in that sector, so from Corollary 1 part (iv) it follows that the average investment of an agent in this economy is $\hat{\boldsymbol{q}} = \frac{1}{\rho}\Sigma^{-1}(\hat{\mathbf{v}} - \hat{w}_0\mathbf{1})$ immediately leading to

$$\frac{d(\widehat{\Delta W})}{d\alpha^n} = \frac{d\hat{w}_0}{d\alpha^n} + \hat{\boldsymbol{q}}^T\left(\frac{d\hat{\mathbf{v}}}{d\alpha^n} - \frac{d\hat{w}_0}{d\alpha^n}\mathbf{1}\right).$$

12. The same argument can be made for Sharpe ratios, which is the focus of Bansal, Fang and Yaron (2006).

Finally, the total size of the economy is M, so that $\hat{\boldsymbol{q}}^T \mathbf{1} = 1$. Now, since $\left(\frac{d\hat{\mathbf{v}}}{d\alpha^n}\right)_j = 0$ for $j \neq n$, it follows that

$$\frac{d(\widehat{\Delta W})}{d\alpha^n} = (\hat{\boldsymbol{q}})_n \frac{d\hat{v}^n}{d\alpha^n}.$$

On inspection, $\frac{d(\widehat{\Delta W})}{d\alpha^n}$ is therefore strictly positive if and only if $\frac{\partial \hat{v}^n}{\partial \alpha^n} > 0$. In the frictionless economy, the condition is $\kappa\gamma < \frac{(\alpha^n)^2}{k^2}$. This is intuitive: if the cost of entering the market is low relative to the productivity, then social welfare increases if there is a productivity shock as capital flows easily into the newly more productive industries.

In the presence of moral hazard, the general form of the change in the certainty equivalent is similar, however, equilibrium values of the market returns and the effort-free wage differ. Specifically, as above we obtain

$$\frac{d(\Delta W)}{d\alpha^n} = (\boldsymbol{q})_n \frac{\partial v^n}{\partial \alpha^n},$$

where $\boldsymbol{q} = \Sigma^{-1}(\mathbf{v} - w_0 \mathbf{1})/\rho$ is the portfolio of a hypothetical worker, working in a risk-free industry. The difference between the two expressions is the adjustment for moral hazard, *i.e.* v^n, rather than $\hat{v}^n$. The condition for a productivity shock to increase equilibrium total productivity under moral hazard ($\partial v^n/\partial \alpha^n > 0$) is somewhat different: $\kappa\gamma < \frac{(b^n)^2(\alpha^n)^3}{k^2}$. Therefore, consider an industry with entry costs in the range

$$\kappa\gamma \in \left(\frac{\alpha^n}{k}\right)^2 [(b^n)^2 \alpha^n, 1].$$

For such an industry, a productivity shock decreases welfare in the economy with moral hazard, whereas in the friction-free economy a shock would increase welfare. Further, as b^n is decreasing in the degree of moral hazard measured by σ_ε, the larger the moral hazard problem the more likely a productivity shock is to lead to a decrease in welfare. Also, for a fixed level of moral hazard, because high-productivity industries have higher incentive pay (b^n is higher), aggregate welfare is more likely to be lower with a positive productivity shock among these firms.

4.7. *Cross-sectional asset pricing implications*

In addition to the real economy, our model can explain well-known asset pricing anomalies and has novel portfolio implications. These results flow from the fact that part of the agents' incentive payment includes systematic risk. So far, we have mainly studied the effects of industry production characteristics, part A in Figure 2, on industry size, cost of capital, portfolio choice, and wages, *i.e.* on parts B and D. In this section, we focus on how A affects equilibrium firm characteristics, part C, which may lead to asset pricing anomalies in the cross section of stock returns.

Our asset pricing implications are all driven by the same intuition: as the production technologies drive different types of wage contracts, portfolios sorted along variables correlated with different technologies should differ systematically from the CAPM. Further, different industries optimally have different types of remuneration contracts and these can be related to stock market returns, as well as to investors' investments. Since firm characteristics—like firm size and book-to-market—in equilibrium will be related to the production technologies, they will also be related to systematic deviations from the CAPM.

High-productivity firms pay out a higher fraction of systematic $\tilde{x}$-risk through wages than low-productivity firms because they gain more if workers exert high effort. That is, even though the firm bears the cost of the idiosyncratic risk imposed on the worker through the incentive contract, it is still optimal to offer such contracts because of the value generated by higher worker effort. In light of this, a partial equilibrium conclusion would be that the firms' returns should be less risky and therefore lower. This is incorrect: even though a sector seems low risk in the market, economy-wide it is not because firms have still paid out a substantial portion of the risk that they have "produced" through the wage bill. Therefore, if an econometrician estimates a CAPM model using the stock market as the market portfolio, the CAPM will fail and firm production characteristics will provide additional explanatory power.

The deviations from the CAPM can be non-linear in our explanatory variables. Thus, the easiest way to test the model is to sort the data on these variables and then run a CAPM using the stock market as the market portfolio. The model has predictions on the relative size of the intercepts across the two sorted portfolios. In all our corollaries, we are considering the results an econometrician should expect who takes the N industry level stocks and sorts them into two portfolios.

First, consider a sort based on labour productivity. This is typically estimated at the industry level by fitting a variant of a Cobb–Douglas production function.[13] Industries are then ranked based on the size of the productivity, with industry 1 having the lowest and industry N having the highest.

Corollary 3. *Suppose that $\partial v_n/\partial\alpha > 0$ so that the marginal product of capital is increasing in α, then*

(i) *firms with high labour productivity α have positive abnormal returns, i.e. there is an $n_0 \leq N$ such that for all $n \geq n_0$, $\bar{\mu}_n > \beta_n\bar{\mu}_{\text{market}}$,*
(ii) *firms with low α have negative abnormal returns, i.e. there is an $n_0 \geq 1$ such that for all $n \leq n_0$, $\bar{\mu}_n < \beta_n\bar{\mu}_{\text{market}}$.*

In both the first-best and the moral hazard economy, a firm's physical investment is *decreasing* in productivity. Alternatively, as the investment determines the size of the firm in this economy, more productive firms are smaller. This general equilibrium effect comes about because of the concave production technology and the fact that the cost of capital in more productive industries is higher. The latter effect is truly general equilibrium: productive technologies spark competition and the subsequent entry of new firms. Total investment in that sector typically increases, leading to a higher industry cost of capital (as risk-averse investors value this investment less).

Since high-productivity firms are small (condition (iii) in Proposition 1), this immediately implies that a sort based on firm size will also provide abnormal returns. Note also that I_1 is the size of firms in the industry with the lowest labour productivity and therefore the industry with the largest firms.[14]

Corollary 4 (Size Effect).

(i) *Small firms have higher expected returns than those predicted by the CAPM, i.e. there is an $I \geq I_N$ such that for all industries with $I_n \leq I$, $\bar{\mu}_n > \beta_n\bar{\mu}_{\text{market}}$.*

13. Kruse (1992) presents such an estimation.

14. In the model, we make specific assumptions on firms' production and cost functions. The corollary, however, does not depend on the specific functional form, as long as the cost is an increasing function of productivity. Any cost function of the form $\kappa + \gamma(\alpha)I^2$, where $\gamma(\cdot)$ is increasing, will lead to a size anomaly.

(ii) *Large firms have lower expected returns than those predicted by the CAPM, i.e. there is an* $I \leq I_1$ *such that for all industries with* $I_n \geq I$, $\bar{\mu}_n < \beta_n \bar{\mu}_{\text{market}}$.

A value-like effect can also arise within the model. The firm raises I in capital, which in the competitive market is just the firm's $t = 1$ market value. After raising this money, it immediately incurs fixed capital costs and pays out the fixed part of wages (see Figure 1). This leads to negative retained earnings, which decreases the book value of assets which are therefore $I_n - \kappa - \gamma \alpha_n I_n^2 - s_n I_n$ at $t = 1$. The market value of the firm, however, does not change. Therefore, the book-to-market ratio in an industry, at $t = 1$, is

$$\text{BM}_n = 1 - \frac{\kappa + \gamma \alpha_n I_n^2 + s_n I_n}{I_n}.$$

The crucial condition that ensures a value effect for all firms is that firms with higher productivity decrease the fixed wage part of workers' compensation more than enough to offset the increase in spending on R & D (which will have an offsetting effect). In this case, high-α firms have high book-to-market ratios and we immediately get the following.

Corollary 5 (Value Effect). *If* $\frac{ds}{d\alpha} < -\sqrt{\frac{\kappa\gamma}{\alpha}}$ *for all* $\alpha \in [\alpha_1, \alpha_N]$, *then*

(i) *firms with high book-to-market ratios have expected returns greater than those predicted by the CAPM, i.e. there is a* BM $\leq$ BM$_N$ *such that for all industries with* BM$_n \geq$ BM, $\bar{\mu}_n > \beta_n \bar{\mu}_{\text{market}}$;

(ii) *firm with low book-to-market ratios have lower expected returns than those predicted by the CAPM, i.e. there is a* BM $\geq$ BM$_1$ *such that for all industries with* BM$_n \leq$ BM, $\bar{\mu}_n < \beta_n \bar{\mu}_{\text{market}}$.

Finally, consider the effect of differing productivity on the fixed part of the contract, s, (through (vii)). For industries with low α_n, $s \approx (1 - b_n) w_0$ and is decreasing and approaches w_0 as α approaches zero. Indeed, as long as $b_n < \frac{w_0}{\rho \sigma_{\varepsilon,n}^2}$, s is decreasing in α. Similarly, for high α $\left(\text{i.e. } \alpha_n > \frac{\rho^2 \sigma_{\varepsilon,n}^4}{4\kappa\gamma}\right)$, s is decreasing in α. Thus, for low-productivity and high-productivity industries, the fixed wage part of the contract is decreasing in productivity. For intermediate values of α, however, the sign of $ds/d\alpha$ is ambiguous.

There is some evidence that the pricing "anomalies" relative to the CAPM depend on both the country and the time period. However, Fama and French (1993) have established strong international evidence for abnormal returns in high book-to-market returns. Our results suggest that national labour conditions should affect returns. For example, in line with Corollary 5 industry-level data on how fixed wages vary with productivity should indicate if we expect a value effect to hold. To illustrate this, we construct a stylized numerical example and run a few standard linear asset pricing regressions.[15] We study an economy with 30 sectors, in which the productivity (α) and moral hazard (σ_ε) vary.

To highlight the effect of differences in labour productivity, we assume that all other parameters are fixed across industries. Specifically, we set $\gamma = \kappa = 1$. We adopt a symmetric structure for the real risks in the economy ($\tilde{x}_n$) risk and let the variance of each of the risk factors be 5, and the covariances between each of the industry risks be 0·01. We fix the mass of agents in the

15. Since our model assumes CARA expected utility, which is needed to make the contracting structure tractable, a full-scale calibration to real variables is not within the scope of our model.

TABLE 1
Summary of parameters in 30-sector economy

$N = 30$	$\alpha_1 = 1{\cdot}61$	$\alpha_{30} = 2{\cdot}25$	$\rho = 2{\cdot}5$
$k = 1$	$\kappa = 1$	$\gamma = 1$	$M = 1$
$\min(\sigma_{\varepsilon,n}) = 1{\cdot}00$	$\max(\sigma_{\varepsilon,n}) = 1{\cdot}03$	$\Sigma_{n,n} = 5$	$\Sigma_{n,j} = 5{\cdot}01$

economy at $M = 1$, each of which has risk aversion coefficient given by $\rho = 2{\cdot}5$. The parameters of the economy are summarized in Table 1.

In this economy, the risk-free rate is zero and so expected returns are also excess returns. A standard CAPM analysis suggests that a regression of expected returns on β should yield the constant market risk premium. In the presence of moral hazard however, variables that are related to the wage bill should explain market returns. In Figure 3, we present three separate regressions of μ, on observed betas, size, and book-to-market, respectively.

Expected return as a function of beta is the non-linear thick solid line market with diamonds. The non-linearity is driven by the cross-sectional differences of wage risk. Therefore, the regression of market beta versus expected return (the thin straight line with stars) does not capture

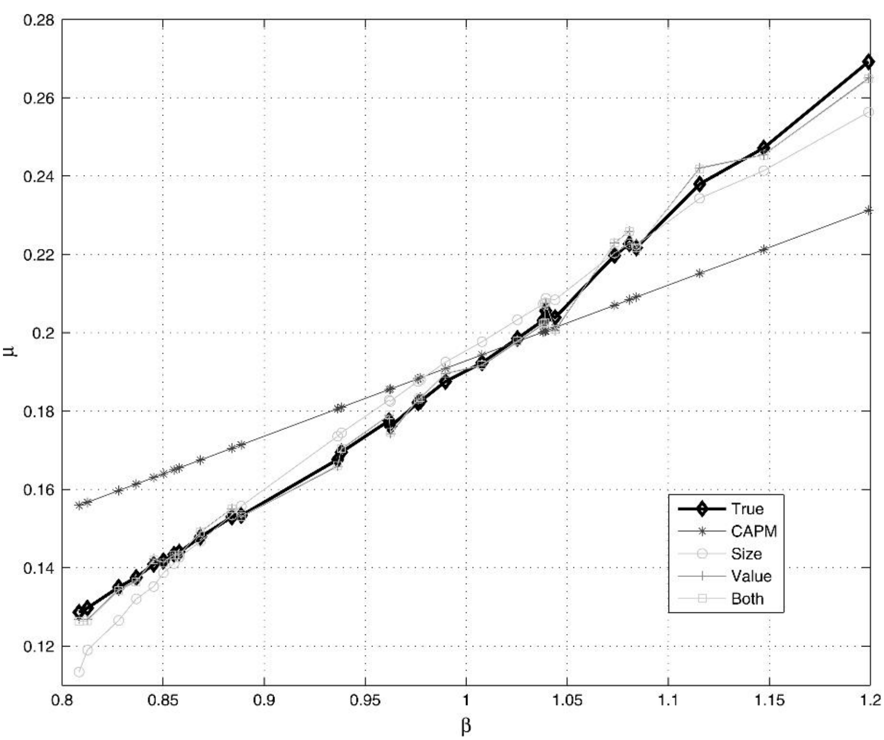

FIGURE 3

Expected returns in five-industry example with moral hazard. The x-axis represents the market β of the stocks and the y-axis their expected returns. True expected returns are represented by the thick line with diamonds. Regressed returns on market beta (thin line with stars), on size (thin line with circles), on book-to-market ratio (thin line with pluses), and on both size and book-to-market (thin line with squares) are also shown. Firm characteristics do a better job at capturing cross-sectional variation in stock returns than market betas expected returns well. By contrast, size (line with circles) and book-to-market (thin line with pluses) do a better job at capturing cross-sectional differences in expected returns. When expected returns are regressed on both size and value (thin line with squares), the result is almost indistinguishable from true expected returns. In line with Corollary 3, there is a threshold above which all firms have positive abnormal returns with respect to the CAPM (these are the high-α firms) and below which all firms have negative abnormal returns (the low-α firms). Thus, in line with our predictions, firm characteristics do a much better job at capturing the cross section of stock returns than systematic risk measured by the CAPM.

We have couched our empirical predictions in terms of portfolios sorted on labour productivity. This is the easiest way of side-stepping any possible non-linearities. However, in a linear regression of the excess return on a stock on the excess return of the market, proxies for the wage bill would be significant. That is, if the market is measured from observations on traded stock returns, then proxies for either labour productivity or the wage bill should have significant explanatory power. Although out of the scope of this paper, such a test would be able to confirm or reject whether our model provides a viable explanation of the cross section of expected returns.

5. CONCLUSIONS

We have characterized a tractable general equilibrium model of production in which workers, remunerated by firms, hedge labour income in financial markets. The CARA framework with normally distributed risks admits both simple, optimal incentive contracts and closed-form solutions for capital market equilibrium.

Understanding the implications of labour supply is important to understand both capital market equilibrium and labour market equilibrium. First, there is empirical evidence that human capital can help explain returns both in the time series and in the cross section.[16] However, there are few theoretical predictions on how the cross section of returns should relate to labour productivity. Second, few labour papers explicitly take into account the fact that workers can trade in financial markets. Finally, partial equilibrium models take firms' cost of capital as exogenous, when clearly it must depend on other investment opportunities.

More broadly, as investors are also workers, firm characteristics should help explain the cross section of returns. Specifically, asset pricing, firm balance sheet characteristics, and returns to human capital are jointly determined in equilibrium. Our analysis leads to several novel empirical implications. It suggests that a firm's expected returns, as well as other firm characteristics—such as size—are directly related to the type of compensation it offers to its workers. Our analysis also relates workers' portfolio decisions to the type of firm they work for.

The model has very specific implications for agents' portfolio holdings. Indeed, given data on agents' portfolios and employment contracts, one could directly test to see if, given an incentive contract, portfolio holdings did accord with the partial equilibrium predictions. The simplest way to do so would be to see if agents change their portfolios when they change sectors. For example, if the wage channel is important, one could observe if a worker moving into a more productive industry will reduce his exposure to financial assets and hold a less diversified portfolio. These actions have immediate risk implications for his portfolio. Specifically, a worker moving into a more productive industry should decrease the stock market exposure of his portfolio if the

16. The seminal time series paper is due to Jagannathan and Wang (1996), while using micro data, Malloy, Moskowitz and Vissing-Jorgenson (2005) find evidence that labour income risk (through a firing decision) can explain the value effect. Further use of human capital proxies has improved the performance of the conditional CAPM, as in, *e.g.* Palacios-Huerta (2003).

industry has returns that are positively correlated with the market and increase the stock market exposure of his portfolio if the industry has returns that are negatively correlated with the market.

APPENDIX

Proof of Lemma 1. The agent's optimization problem (4) takes the form

$$\begin{aligned}\Delta W_n &= \max_{e_n,\boldsymbol{q}_n} \upsilon_n \\ &= \max_{e_n,\boldsymbol{q}_n}\Big[s_n + b_n\alpha_n e_n + \bar{\boldsymbol{\mu}}^T\boldsymbol{q}_n - \frac{k}{2}e_n^2 - \\ &\quad -\frac{\rho}{2}(b_n^2\sigma_{x,n}^2 + b_n^2\sigma_{\varepsilon,n}^2 + (\boldsymbol{q}_n)^T\Sigma_\mu(\boldsymbol{q}_n) + 2b_n\boldsymbol{\sigma}_{\mu,n}^T\boldsymbol{q}_n)\Big].\end{aligned}$$

The first-order conditions are therefore

$$\frac{\partial \upsilon_n}{\partial e_n} : b_n\alpha_n - ke_n = 0 \Rightarrow e_n = \frac{\alpha_n b_n}{k}, \tag{A.1}$$

$$\frac{\partial \upsilon_n}{\partial \boldsymbol{q}_n} : \bar{\boldsymbol{\mu}} - \rho(\Sigma_\mu\boldsymbol{q}_n + b_n\boldsymbol{\sigma}_{\mu,n}) = \boldsymbol{0} \Rightarrow \boldsymbol{q}_n = \Sigma_\mu^{-1}\left(\frac{\bar{\boldsymbol{\mu}}}{\rho} - b_n\boldsymbol{\sigma}_{\mu,n}\right). \tag{A.2}$$

As the decision variables are separated over e_n and $\boldsymbol{q}_n$ and as the highest order terms in e_n and $\boldsymbol{q}_n$ are strictly negative-definite quadratic forms, a solution to the first-order conditions is also a global maximum. ‖

Proof of Lemma 2. Equation (A.2) implies the following values:

$$\begin{aligned}\bar{\boldsymbol{\mu}}^T\boldsymbol{q}_n &= \frac{A}{\rho} - b_n B_n, \\ \boldsymbol{q}_n^T\Sigma_\mu(\boldsymbol{q}_n) &= \frac{A}{\rho^2} + b_n^2 C_n - \frac{2b_n b_n}{\rho}, \\ \boldsymbol{\sigma}_{\mu,n}^T\boldsymbol{q}^n &= \frac{b_n}{\rho} - b_n C_n.\end{aligned}$$

If the participation constraint of a worker in sector n binds, then

$$\begin{aligned}\Delta W &= s_n + b_n\bar{x}_n + b_n\alpha_n\frac{\alpha_n b_n}{k} + \left(\frac{A}{\rho} - b_n B_n\right) - \frac{(\alpha_n b_n)^2}{2k} \\ &\quad -\frac{\rho}{2}\left(b_n^2\sigma_{x,n}^2 + b_n^2\sigma_{\varepsilon,n}^2 + \frac{A}{\rho^2} + b_n^2 C_n - \frac{2b_n B_n}{\rho} + 2b_n\left(\frac{B_n}{\rho} - b_n C_n\right)\right) \\ &= s_n + b_n^2\left(\frac{\alpha_n^2}{2k} + \frac{\rho}{2}(C_n - \sigma_{x,n}^2) - \frac{\rho}{2}\sigma_{\varepsilon,n}^2\right) - b_n(B_n - \bar{x}_n) + \frac{A}{2\rho}.\end{aligned}$$

Therefore, the fixed wage at which the worker is indifferent between accepting employment in industry n and going elsewhere is

$$s_n(b_n, I_n) = b_n^2\left(-\frac{\alpha_n^2}{2k} - \frac{\rho}{2}(C_n - \sigma_{x,n}^2) + \frac{\rho}{2}\sigma_{\varepsilon,n}^2\right) + b_n(B_n - \bar{x}_n) + \Delta W - \frac{A}{2\rho}. \quad ‖$$

Proof of Lemma 3. From equation (7), the profit function, a firm in sector n solves

$$\begin{aligned}\max_{I_n,b_n}\bar{\pi}_n &= E[\tilde{\pi}_n] = I_n((\alpha_n e_n + \bar{x}_n)(1-b_n) - s_n(b_n, I_n) - (1-b_n)z_n - \gamma_n I_n) - \kappa_n \\ &= \max_{I_n,b_n} I_n\left(\left(\frac{b_n\alpha_n^2}{k} + \bar{x}_n\right)(1-b_n) - s_n(b_n, I_n) - (1-b_n)z_n - \gamma_n I_n\right) - \kappa_n\end{aligned}$$

$$= \max_{I_n, b_n} I_n \left(\left(\left(\frac{b_n \alpha_n^2}{k} + \bar{x}_n \right) (1-b_n) - \left(b_n^2 \left(-\frac{\alpha_n^2}{2k} - \frac{\rho}{2}(C_n - \sigma_{x,n}^2) \right.\right.\right.\right.$$

$$\left.\left.\left. + \frac{\rho}{2}\sigma_{\varepsilon,n}^2 \right) + b_n(B_n - \bar{x}_n) + \Delta W - \frac{A}{2\rho} \right) - (1-b_n) z_n - \gamma_n I_n \right) - \kappa_n$$

$$= \max_{I_n, b_n} I_n \left(b_n^2 \left(-\frac{\alpha_n^2}{2k} + \frac{\rho}{2}(C_n - \sigma_{x,n}^2) - \frac{\rho}{2}\sigma_{\varepsilon,n}^2 \right) \right.$$

$$\left. + b_n \left(\frac{\alpha_n^2}{k} + z_n - b_n \right) + \frac{A}{2\rho} - \Delta W + \bar{x}_n - z_n \right) - \alpha_n \gamma_n I_n^2 - \kappa_n$$

$$\stackrel{\text{def}}{=} I_n (M_2^n b_n^2 + M_1^n b_n + M_0^n) - \alpha_n \gamma_n I_n^2 - \kappa_n.$$

The first-order condition in b_n is

$$\frac{\partial \bar{\pi}_n}{\partial b_n} : 2M_2^n b_n + M_1^n = 0$$

$$\Rightarrow\ b_n = -\frac{M_1^n}{2M_2^n} = \frac{\alpha_n^2/k + z_n - B_n}{\alpha_n^2/k + \rho(\sigma_{x,n}^2 - C_n + \sigma_{\varepsilon,n}^2)}. \tag{A.3}$$

From the Cauchy–Schwarz inequality, $\text{cov}(\tilde{x}_n, \tilde{y}) \le \sqrt{\text{var}(\tilde{x}_n)\text{var}(\tilde{y})}$. Therefore, $C_n \le \sigma_{x,n}^2$: For the choice $\tilde{y} = \mathbf{a}^T \tilde{\boldsymbol{\mu}}$ with $\mathbf{a} = \Sigma_\mu^{-1} \boldsymbol{\sigma}_{\mu,n}$, the inequality leads to $\boldsymbol{\sigma}_{\mu,n}^T \Sigma_\mu^{-1} \boldsymbol{\sigma}_{\mu,n} \le \sqrt{\boldsymbol{\sigma}_{\mu n}^T \Sigma_\mu^{-1} \boldsymbol{\sigma}_{\mu,n} \times \sigma_{x,n}^2}$, *i.e.* $C_n \le \sqrt{C_n} \times \sqrt{\sigma_{x,n}^2}$. This immediately implies that M_2 is strictly negative, and in fact, $M_2 < -\alpha_n^2/k$.

The first-order condition for I_n is

$$\frac{\partial \bar{\pi}}{\partial I} : M_2^n b_n^2 + M_1^n b_n + M_0^n - 2\gamma_n I_n = 0 \Longrightarrow I_n = \frac{M_2^n b_n^2 + M_1^n b_n + M_0^n}{2\alpha_n \gamma_n}. \tag{A.4}$$

From the first order condition (f.o.c.) on b_n, equation (A.3), this is equivalent to

$$I_n = \frac{1}{2\alpha_n \gamma_n} \left(-\frac{(M_1^n)^2}{4M_2^n} + M_0^n \right) \stackrel{\text{def}}{=} \frac{T^n}{2\alpha_n \gamma_n}. \tag{A.5}$$

We note that solution, (b_n, I_n), to the first-order condition is unique. We now prove that it is a global maximum as long as $I_n > 0$ and $0 < b_n < 1$. To see this, substitute the solutions to the f.o.c. into the profit function and obtain

$$\bar{\pi} = \frac{(T^n)^2}{4\alpha_n \gamma_n} - \kappa_n$$

$$= \frac{1}{4\alpha_n \gamma_n} \left(\frac{(\alpha_n^2/k + z_n - B_n)^2}{\alpha_n^2/k + \rho(\sigma_{x,n}^2 - C_n + \sigma_{\varepsilon,n}^2)} + \frac{A}{2\rho} - \Delta W + \bar{x}_n - z_n \right)^2 - \kappa_n \ge -\kappa_n.$$

Therefore, any strictly better strategy must lead to a profit greater than $-\kappa_n$.

We establish that no other solution yields such a profit. Clearly, the optimization problem is smooth, so an optimum will either be at a boundary (including the extended boundary $I_n = \infty$) or satisfy the first-order conditions. As the solution to the f.o.c. is unique, we check the boundaries. Suppose that $I_n = \infty$: the Hessian of this optimization is of the form $H = [2M_2^n I_n, M_1^n; M_1^n, -2\gamma_n]$, with characteristic equation $(\lambda - 2M_2^n I_n)(\lambda + 2\gamma_n) - (M_1^n)^2 = 0$, *i.e.* $\lambda^2 + 2(\gamma_n - I_n M_2^n)\lambda - 4\gamma_n I_n M_2^n - (M_1^n)^2 \stackrel{\text{def}}{=} \lambda^2 + a_1 \lambda + a_0 = 0$. Clearly, $a_1 > 0$, $a_0 < a_1$, and, for large enough I_n, $a_0 > 0$, which implies that, for large enough I_n, both characteristic roots are negative. Thus, there is always an I_n such that $\bar{\pi}$ is decreasing regardless of b_n and the optimum cannot be reached at the (extended) boundary $I_n = \infty$.

Moreover, any $b_n \ge 1$ will lead to $\bar{\pi} \le -\kappa$, so no interesting solution can have $b_n \ge 1$ and the boundary $I_n = 0$ will lead to $\bar{\pi} = -\kappa$. Thus, any non-interior optimum must lie on the boundary $b_n = 0$. On this boundary, the optimal investment level is $I_n = M_0^n / 2\gamma_n$, which is feasible if $M_0^n \ge 0$ (as otherwise $I_n < 0$). In this case, $\bar{\pi} = M_0^2 / 4\gamma_n - \kappa_n$. However, if $M_0^n \ge 0$, then this boundary solution is obviously dominated by $\bar{\pi} = (-(M_1^n)^2 / 4M_2^n + M_0)^2 / 4\gamma_n - \kappa_n$ (as $M_2^n < 0$), so no solution on the boundary $b_n = 0$ can dominate the interior solution. ‖

Proof of Lemma 4.

(i) With zero expected economic profits, the expected return of a share in industry n is

$$\mu_n = \frac{E[S_n(2)] - S_n(1)}{S_n(1)} = \frac{E[\tilde{\pi}_{n,\ell}] + r_n I_{n,\ell}}{I_{n,\ell}} = \frac{0 + r_n I_{n,\ell}}{I_{n,\ell}} = r_n.$$

(ii) Since $\tilde{\mu}_n = \frac{\tilde{\pi}_{n,\ell}}{I_{n,\ell}} + r_n = c_n + (1-b_n)\tilde{x}_n$, for some constant c_n, it follows that $\text{cov}(\tilde{\mu}_i, \tilde{\mu}_j) = (1-b_i)\sigma_{i,j}(1-b_j)$. A similar argument holds for $\text{cov}(\tilde{\mu}_i, \tilde{x}_j)$. ‖

Proof of Proposition 1. We first assume that all wage contracts are linear, construct an equilibrium satisfying (i)–(vii), and then show that it is unique. We then show that it is indeed optimal for a firm to offer a linear wage contract, given that all other firms offer linear wage contracts, which completes the proof.

We define $\Lambda_I = \text{diag}(I^1, \ldots, I_N)$, $\Lambda_\alpha = \text{diag}(\alpha_1, \ldots, \alpha_N)$, $\Lambda_b = \text{diag}(b^1, \ldots, b_n)$, and $\Lambda_{1-b} = \text{diag}(1-b^1, \ldots, 1-b_n)$.

First, from Lemma 1, an equilibrium with optimizing workers will satisfy (i).

We note that equation (9) implies that in equilibrium, $C_n = [\Sigma_x]_{n,n} = \sigma^2_{x,n}$ and that $b_n = \bar{\mu}^n/(1-b_n) = z_n$, which in turn, through the relation $r^n = \bar{\mu}^n$, implies (iii):

$$b_n = \frac{1}{1 + \frac{k\rho\sigma^2_{\varepsilon,n}}{\alpha_n^2}} \in (0,1).$$

Also, $M_1^n = \alpha_n^2/k$ and $M_2^n = -(\alpha_n^2/k + \rho\sigma^2_{\varepsilon,n})/2$, where M_1^n and M_2^n were defined in the proof of Lemma 3.

Moreover, equation (8) together with $E[\tilde{\pi}^n] = 0$ implies (ii):

$$I_n = \sqrt{\frac{\kappa_n}{\gamma_n \alpha_n}}, \tag{A.6}$$

which is strictly positive. However, we also have through equation (A.5)

$$2\sqrt{\gamma_n \kappa_n \alpha_n} + \frac{(M_1^n)^2}{4M_2^n} = \frac{A}{2\rho} - \Delta W - z_n, \tag{A.7}$$

which through the relations $\Delta W = w_0 + A/2\rho$ and $\bar{\mu}^n/(1-b_n) = z^n$ leads to

$$\bar{\mu}^n = (1-b_n)\left(-2\sqrt{\gamma_n \kappa_n \alpha_n} - \frac{(M_1^n)^2}{4M_2^n} - w_0\right) \Rightarrow \bar{\boldsymbol{\mu}} = \Lambda_{1-b}(\boldsymbol{v} - w_0 \mathbf{1}). \tag{A.8}$$

The market-clearing condition in the stock market now gives us

$$\boldsymbol{Q}\Lambda_I \boldsymbol{L} = \Lambda_I \boldsymbol{L},$$

which implies that

$$(\bar{\boldsymbol{I}} - \boldsymbol{Q})\Lambda_I \boldsymbol{L} = \mathbf{0} \tag{A.9}$$

(where $\bar{\boldsymbol{I}}$ is the identity matrix), *i.e.* 1 must be an eigenvalue to $\boldsymbol{Q}$, with eigenvector $\boldsymbol{\lambda}$. Given such a $\boldsymbol{\lambda}$, the vector of firm mass $\boldsymbol{L} = \frac{M}{\mathbf{1}^T\boldsymbol{\lambda}}\Lambda_I^{-1}\boldsymbol{\lambda}$ will be a solution. Moreover, the labor market condition, $M = \boldsymbol{I}^T\boldsymbol{L}$, can be rewritten as

$$\mathbf{1}^T \Lambda_I \boldsymbol{L} = M. \tag{A.10}$$

We have

$$\boldsymbol{Q} = \Sigma_\mu^{-1}\left(\frac{\bar{\boldsymbol{\mu}}\mathbf{1}^T}{\rho} - \Sigma_{\mu,x}\Lambda_b\right), \tag{A.11}$$

where $\Sigma_{\mu,x} = [\boldsymbol{\sigma}_{\mu,1}, \ldots, \boldsymbol{\sigma}_{\mu,n}]$. Since from equation (9), we know that $\Sigma_\mu = (\bar{\boldsymbol{I}} - \Lambda_b)\Sigma(\bar{\boldsymbol{I}} - \Lambda_b)$ and $\Sigma_{\mu,x} = (\bar{\boldsymbol{I}} - \Lambda_b)\Sigma$, equation (A.11) can be rewritten as

$$\boldsymbol{Q} = (\bar{\boldsymbol{I}} - \Lambda_b)^{-1}\Sigma^{-1}(\bar{\boldsymbol{I}} - \Lambda_b)^{-1}\left(\frac{\bar{\boldsymbol{\mu}}\mathbf{1}^T}{\rho} - (\bar{\boldsymbol{I}} - \Lambda_b)\Sigma\Lambda_b\right).$$

This in turn, via equations (A.8) and (A.11), means that equation (A.9) can be rewritten as

$$\begin{aligned}\Sigma_\mu^{-1}\left(\frac{\bar{\boldsymbol{\mu}}\mathbf{1}^T}{\rho}-\Sigma_{\mu,x}\Lambda_b\right)\boldsymbol{\lambda} &= \boldsymbol{\lambda}\\ &= \Sigma_\mu^{-1}\left(\frac{\Lambda_{1-b}(\boldsymbol{v}-w_0\mathbf{1})\mathbf{1}^T}{\rho}-\Sigma_{\mu,x}\Lambda_b\right)\boldsymbol{\lambda}\\ &= \Sigma_\mu^{-1}\left(\frac{\Lambda_{1-b}\boldsymbol{v}\mathbf{1}^T}{\rho}-\Sigma_{\mu,x}\Lambda_b\right)\boldsymbol{\lambda}-\frac{w_0 M}{\rho}\Sigma_\mu^{-1}\Lambda_{1-b}\mathbf{1}\\ &\Rightarrow \frac{Mw_0}{\rho}(\Sigma_\mu^{-1}(\Lambda_{1-b}\mathbf{v}\mathbf{1}^T/\rho-\Sigma_{\mu,x}\Lambda_b)-\bar{\boldsymbol{I}})^{-1}\Sigma_\mu^{-1}\Lambda_{1-b}\mathbf{1}\\ &= \boldsymbol{\lambda},\end{aligned}$$

so

$$\boldsymbol{\lambda}=\frac{Mw_0}{\rho}\boldsymbol{Z}\mathbf{1},$$

where $\boldsymbol{Z}=(\Sigma_\mu^{-1}(\Lambda_{1-b}\mathbf{v}\mathbf{1}^T/\rho-\Sigma_{\mu,x}\Lambda_b)-\bar{\boldsymbol{I}})^{-1}\Sigma_\mu^{-1}\Lambda_{1-b}$. Since $\mathbf{1}^T\lambda=M$, we have

$$w_0=\frac{\rho}{\mathbf{1}^T\boldsymbol{Z}\mathbf{1}}.$$

Now, expanding $\boldsymbol{Z}$ through the definitions of Σ_μ and $\Sigma_{\mu,x}$ leads to

$$\boldsymbol{Z}=\left(\frac{\mathbf{v}\mathbf{1}^T}{\rho}-\Sigma\right)^{-1},$$

which through the Sherman–Morrison–Woodberg formula,

$$(\boldsymbol{A}+\boldsymbol{U}\boldsymbol{V}^T)^{-1}=\boldsymbol{A}^{-1}-\boldsymbol{A}^{-1}\boldsymbol{U}(I+\boldsymbol{V}^T\boldsymbol{A}^{-1}\boldsymbol{U})^{-1}\boldsymbol{V}^T\boldsymbol{A}^{-1},$$

leads to (iv): $w_0=\rho/(\mathbf{1}^T\boldsymbol{Z}\mathbf{1})=\frac{\mathbf{1}^T\Sigma^{-1}\mathbf{v}-\rho}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}}$.

From equation (A.12) and the relation $\boldsymbol{L}=\frac{M}{\mathbf{1}^T\boldsymbol{\lambda}}\Lambda_I^{-1}\boldsymbol{\lambda}$, it follows that

$$\boldsymbol{L}=Mw_0\rho\Lambda_I^{-1}\mathbf{Z}\mathbf{1},$$

and another application of the Sherman–Morrison–Woodberg formula finally leads to (v).

Condition (vi), $W+w_0+(\boldsymbol{v}-w_0\mathbf{1})^T\Sigma^{-1}(\boldsymbol{v}-w_0\mathbf{1})/2\rho$, follows from the relation $\Delta W=w_0+A/2\rho$ and equation (A.8).

Condition (vii) follows immediately from plugging the derived values of variables into equation (7) of Lemma (2).

Under Assumption 1 part (i), (iii) implies that w_0 is strictly positive. Moreover, from (vi), it follows that

$$\Lambda_I\boldsymbol{L}=c\Sigma^{-1}(\boldsymbol{v}-w_0\mathbf{1}),$$

where $c>0$, and through (v) it then follows that $\boldsymbol{L}>\mathbf{0}$ if and only if

$$(\mathbf{1}\Sigma^{-1}\mathbf{1})\Sigma^{-1}\boldsymbol{v}-(\mathbf{1}\Sigma^{-1}\boldsymbol{v})\Sigma^{-1}\mathbf{1}+\rho\Sigma^{-1}\mathbf{1}>\mathbf{0},$$

which in turn is guaranteed by Assumption1 part (ii). Thus, all variables are strictly positive so the equilibrium is interior. Moreover, each of equations (A.6)–(A.10) is necessary, so the equilibrium $\mathcal{X}=(w_0,\boldsymbol{L},\boldsymbol{I},\boldsymbol{e},\bar{\boldsymbol{\mu}},\mathbf{q})$ is unique.

Optimality of the linear contract

It remains to be shown that it is optimal for a firm to offer a linear compensation contract, given that all other firms offer linear contracts. Given that all but a zero measure of firms offer linear contracts, then it immediately follows that, at $t=1$, the price in the market of any asset with pay-off $\tilde{y}$ at $t=2$ is

$$P[\tilde{X}]=E[\tilde{M}\tilde{X}], \tag{A.12}$$

where the pricing kernel $\tilde{M} = Ce^{-\rho \boldsymbol{q}^T \tilde{\boldsymbol{z}}} = C' e^{-\rho \boldsymbol{q}^T \Lambda_{1+\mathbf{z}} \tilde{\boldsymbol{y}}}$, $\boldsymbol{q} = \frac{1}{\rho} \Sigma_z^{-1} \mathbf{z}$, $\mathbf{z} = \Lambda_{\boldsymbol{b}}^{-1} \bar{\boldsymbol{\mu}}$, $\tilde{\boldsymbol{z}} = \Lambda_{\mathbf{1}+\mathbf{z}} \tilde{\boldsymbol{x}} - \mathbf{1}$, $\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{x}} - \mathbf{1}$, $\Sigma_Z = \Lambda_{1+\mathbf{z}} \Sigma \Lambda_{1+\mathbf{z}}$, and C' is chosen such that $E[\tilde{M}] = 1$ (to price the risk-free asset correctly). Since

$$1 = C' E\left[e^{-\rho \boldsymbol{q}^T \Lambda_{1+\mathbf{z}} \tilde{\boldsymbol{y}}}\right] \tag{A.13}$$

$$= C' \frac{1}{\sqrt{(2\pi)^N \operatorname{Det}\Sigma}} \int_{\mathbb{R}^N} e^{-\rho \boldsymbol{q}^T \boldsymbol{y} - \frac{1}{2} \boldsymbol{y}^T \Sigma \boldsymbol{y}} d\boldsymbol{y} \tag{A.14}$$

$$= C' \frac{1}{\sqrt{(2\pi)^N \operatorname{Det}\Sigma}} \int_{\mathbb{R}^N} e^{-\frac{1}{2}(\boldsymbol{y}+\rho\Sigma\Lambda_{1+\mathbf{z}}\boldsymbol{q})^T \Sigma^{-1}(\boldsymbol{y}+\rho\Sigma\Lambda_{1+\mathbf{z}}\boldsymbol{q}) + \frac{\rho^2}{2}\boldsymbol{q}^T \Lambda_{1+\mathbf{z}}\Sigma\Lambda_{1+\mathbf{z}}\boldsymbol{q}} d\boldsymbol{y} \tag{A.15}$$

$$= C' \frac{1}{\sqrt{(2\pi)^N \operatorname{Det}\Sigma}} \int_{\mathbb{R}^N} e^{-\frac{1}{2}(\boldsymbol{y}+\Lambda_{1+\mathbf{z}}^{-1}\mathbf{z})^T \Sigma^{-1}(\boldsymbol{y}+\Lambda_{1+\mathbf{z}}^{-1}\mathbf{z}) + \frac{1}{2}\mathbf{z}^T \Sigma_z^{-1}\mathbf{z}} d\boldsymbol{y} \tag{A.16}$$

$$= C' e^{\frac{1}{2}\mathbf{z}^T \Sigma_z^{-1} \mathbf{z}}, \tag{A.17}$$

it follows that $C' = e^{-\frac{1}{2}\mathbf{z}^T \Sigma_z^{-1}\mathbf{z}}$. It is easy to show that equation (A.12) reduces to $P[\tilde{\boldsymbol{x}}] = \Lambda_{\mathbf{1}+\mathbf{z}}^{-1}\mathbf{1}$, in line with our pricing formula for $\tilde{\boldsymbol{x}}$ risk, but the formula also holds for non-linear risks and defines the firm's measure of general economic profits.

We use the relationship $\mathbf{z} = \Lambda_{\mathbf{1}+\mathbf{z}}\mathbf{1} - \mathbf{1}$ to get $\mathbf{1} - \Lambda_{\mathbf{1}+\mathbf{z}}^{-1}\mathbf{z} = \Lambda_{\mathbf{1}+\mathbf{z}}^{-1}\mathbf{1}$. This allows us to rewrite the pricing equation as

$$P[\tilde{X}] = \frac{1}{\sqrt{(2\pi)^N \operatorname{Det}\Sigma}} \int_{\mathbb{R}^N} e^{-\frac{1}{2}(\boldsymbol{x}-(\mathbf{1}-\Lambda_{\mathbf{1}+\mathbf{z}}^{-1}\mathbf{z}))^T \Sigma^{-1}(\boldsymbol{x}-(\mathbf{1}-\Lambda_{\mathbf{1}+\mathbf{z}}^{-1}\mathbf{z}))} d\boldsymbol{x}$$

$$= \frac{1}{\sqrt{(2\pi)^N \operatorname{Det}\Sigma}} \int_{\mathbb{R}^N} e^{-\frac{1}{2}(\boldsymbol{x}-\Lambda_{\mathbf{1}+\mathbf{z}}^{-1}\mathbf{1})^T \Sigma^{-1}(\boldsymbol{x}-\Lambda_{\mathbf{1}+\mathbf{z}}^{-1}\mathbf{1})} d\boldsymbol{x},$$

which in turn takes the "risk-neutral" form

$$P[\tilde{X}] = E_{\mathbb{Q}}[\tilde{X}], \tag{A.18}$$

where $\tilde{\boldsymbol{x}}$ is $N(\Lambda_{\mathbf{1}+\mathbf{z}}^{-1}\mathbf{1}, \Sigma)$ distributed in the $\mathbb{Q}$ measure. Since $\tilde{\boldsymbol{z}} = \Lambda_{\mathbf{1}+\mathbf{z}}\tilde{\boldsymbol{x}} - \mathbf{1}$, $\tilde{\boldsymbol{z}} \sim N(\mathbf{0}, \Sigma_Z)$ in the $\mathbb{Q}$ measure. The objective of a representative firm in industry n is therefore to maximize

$$P[\tilde{\pi}] = E_{\mathbb{Q}}\left[\alpha I \int_1^2 e(t)dt + \tilde{x}_n I - I \int_1^2 \tilde{\omega}(t)dt - (\kappa + \gamma \alpha_n I_{n,l}^2)\right],$$

and, given the optimal I, which is positive[17], this is realized by optimizing

$$E_{\mathbb{Q}}\left[\alpha \int_1^2 e(t)dt - \int_1^2 \tilde{\omega} dt\right]. \tag{A.19}$$

Under the $\mathbb{Q}$-measure, the firm's optimization problem thus has exactly the same form as in Holmström and Milgrom (1987), in their risk-neutral case.

Now, given that an agent holds the portfolio $\boldsymbol{q}+\boldsymbol{v}$ for some vector $\mathbf{v}$, the agent's expected utility is

$$U = Eu = E\left[-e^{-\rho\left((\boldsymbol{q}+\boldsymbol{v})^T\tilde{\mathbf{z}} - \int_1^2 \frac{e^2(t)}{k}dt + \int_1^2 \tilde{\omega}(t)dt\right)}\right], \tag{A.20}$$

and along identical lines as when rewriting the pricing formula in $\mathbb{Q}$-measure form, this can be rewritten as

$$U = \frac{1}{C'} E_{\mathbb{Q}}\left[-e^{-\rho\left(\boldsymbol{v}^T\tilde{\mathbf{z}} - \int_1^2 \frac{e^2(t)}{k}dt + \int_1^2 \tilde{\omega}(t)dt\right)}\right],$$

i.e.

$$U = \frac{1}{C'} E_{\mathbb{Q}} u\left[\boldsymbol{v}^T\tilde{\mathbf{z}} - \int_1^2 \frac{e^2(t)}{k}dt + \int_1^2 \tilde{\omega}(t)dt\right]. \tag{A.21}$$

The agent will therefore maximize equation (A.21) over $\mathbf{v}$ and effort $e(t)$.

17. It can never be optimal for the firm to choose $I = 0$ since this would lead to strictly negative economic profits of $-\kappa$ and the linear strategy described previously in this proof leads to zero economic profits

First observe that if $\int_1^2 \tilde{w}(t)dt$ is independent of $\tilde{\boldsymbol{x}}$-risk—*e.g.* if it only contains $\tilde{\varepsilon}_n$ risk—then since $\tilde{\boldsymbol{z}}$ has an expectation of zero, the agent will choose $\mathbf{v} = \mathbf{0}$ since any other choice will just introduce noise without returns and the agents utility function will, in this case, also take the same form as in Holmström and Milgrom (1987). The firm's optimization problem of contract over equation (A.19) w.r.t. the agent optimizing equation (A.21) will therefore in this case lead to a linear contract being optimal, following Theorem 7 in Holmström and Milgrom (1987).

Thus, if the firm observes $f(t) \stackrel{\text{def}}{=} \alpha \int_1^t e(s)ds + \varepsilon(t)$, the optimal compensation contract is of the form $w = a_0 + a_1 f(2)$. However, by assumption such a contract is not feasible for the firm to offer because it does not observe $f(t)$. Instead, it observes $g(t) = f(t) + \tilde{x}_n(t) = f(t) + \frac{1}{1+z_n} + \frac{1}{1+z_n}\tilde{z}_n$ (since $g(t) = \tilde{R}/I$). The unobservability of $f(t)$ forces the firm to include $\tilde{z}_n$ risk in the compensation contract. Along the lines of the hedging argument we just made, since the worker can hedge this risk in the market by changing $(\mathbf{v})_n$, the worker is indifferent to promised extra $\tilde{z}_n$ risk. In other words, the worker agrees with the firm that the price for $\tilde{z}_n$ risk is zero. Therefore, any compensation contract of the form $w = a\tilde{z}_n + F(\{f(s)\}, 1 \le s \le 2)$, for $a \neq 0$, will lead to an identical outcome as if $a = 0$—for both the agent and the firm. Therefore, even though contract $w = a_0 + a_1 f(2)$ is not feasible, the firm can implement the equivalent feasible contract $w' = a_0 - a_1 \frac{1}{1+z_n} + a_1 g(t)$, which in line with this argument leads to the same outcome (effort level, expected utility, and economic profits—and it is the only feasible contract that achieves this) as w. This is therefore the uniquely optimal feasible contract. ‖

Proof of Proposition 2. The first part follows immediately from equation (A.8) and $M_1^n = \alpha_n^2/k$, $M_2^n = -(\alpha_n^2/k + \rho\sigma_{\varepsilon,n}^2)/2$.

The value-weighted market portfolio is $\boldsymbol{q} = \text{diag}(\boldsymbol{I})\boldsymbol{L}/(\boldsymbol{I}^T\boldsymbol{L}) = \Sigma^{-1}(\mathbf{v} - w_0\mathbf{1})/\mathbf{1}^T\Sigma^{-1}(\mathbf{v} - w_0\mathbf{1})$. However, the mean–variance efficient portfolio in the financial market is $\boldsymbol{q}^* = \Sigma_\mu^{-1}\bar{\boldsymbol{\mu}}/\mathbf{1}^T\Sigma_\mu^{-1}\bar{\boldsymbol{\mu}} = (\bar{\boldsymbol{I}} - \Lambda_b)^{-1}\Sigma^{-1}(\mathbf{v} - w_0\mathbf{1})/[\mathbf{1}^T(\bar{\boldsymbol{I}} - \Lambda_b)^{-1}\Sigma^{-1}(\mathbf{v} - w_0\mathbf{1})]$. The CAPM will hold with respect to this portfolio (see, *e.g.* Ingersoll 1987) and obviously to any scaled version of this portfolio, *e.g.*

$$\boldsymbol{\nu} = (\bar{\boldsymbol{I}} - \Lambda_b)^{-1}\boldsymbol{q}.$$

Now,

$$b_n = \frac{1}{1 + \frac{k\rho\sigma_{\varepsilon,n}^2}{\alpha_n^2}} \in (0, 1)$$

and since

$$(\bar{\boldsymbol{I}} - \Lambda_b)^{-1} = \text{diag}(1/(1 - b^1), \ldots, 1/(1 - b_n))$$

and also

$$\frac{1}{1 - b_n} = 1 + \frac{\alpha_n}{k\rho\sigma_{\varepsilon,n}^2},$$

this leads to

$$\boldsymbol{\nu} = (I + k^{-1}\rho^{-1}\Lambda^2)\boldsymbol{q}.$$

Since the CAPM holds with respect to $\boldsymbol{\nu}$, we have $\boldsymbol{\mu} = \boldsymbol{\beta}[\boldsymbol{\nu}^T\bar{\boldsymbol{\mu}}]$, where $\boldsymbol{\beta} = \Sigma_\mu\boldsymbol{\nu}/\boldsymbol{\nu}^T\Sigma_\mu\boldsymbol{\nu}$, which, plugging in the definition of $\boldsymbol{\nu}$, leads to equations (12) and (13). ‖

Proof of Proposition 3. Follows directly from Proposition 1 part (v) and Corollary 1 part (iii). ‖

Proof of Proposition 4. Follows directly from Proposition 1 part (vi) and Corollary 1 part (iv). ‖

Proof of Proposition 5. From the discussion in the main text, it is sufficient to show that $[\boldsymbol{X}]_{n,n} > 0$ for $n = 1, \ldots, N$, $N \ge 2$. We have

$$[\boldsymbol{X}]_{n,n} = \delta_n^T\left(\bar{\boldsymbol{I}} - \frac{\Sigma^{-1}\mathbf{1}\mathbf{1}^T}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}}\right)\Sigma^{-1}\delta_n$$
$$= \frac{1}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}}((\delta_n^T\Sigma^{-1}\delta_n)(\mathbf{1}^T\Sigma^{-1}\mathbf{1}) - (\delta_n^T\Sigma^{-1}\mathbf{1})^2),$$

where δ_n is a vector with zeros, except for $[\delta_n]_n = 1$. Since Σ is a symmetric, strictly positive-definite matrix, so is Σ^{-1}. Therefore, the inner product $\langle \boldsymbol{a}, \boldsymbol{b}\rangle = \boldsymbol{a}^T\Sigma^{-1}\boldsymbol{b}$ and the norm $\|\boldsymbol{a}\|^2 = \langle \boldsymbol{a}, \boldsymbol{a}\rangle$ can be defined.

From Cauchy–Schwarz inequality, it follows that $\langle \delta_n, \mathbf{1}\rangle < \|\delta_n\|\|\mathbf{1}\|$, $N \ge 2$, which immediately implies that $(\delta_n^T\Sigma^{-1}\delta_n)(\mathbf{1}^T\Sigma^{-1}\mathbf{1}) - (\delta_n^T\Sigma^{-1}\mathbf{1})^2 > 0$, so the result follows. ‖

Proof of Lemma 5. From Proposition 1, it follows that $\partial r_n/\partial\alpha_p = \partial v_n/\partial\alpha_p - \partial w_0/\partial\alpha_p$. Since $\Sigma = c_0\bar{\boldsymbol{I}} + c_1\mathbf{1}\mathbf{1}^T$, it follows that

$$\Sigma^{-1} = \frac{1}{c_0}\left(\bar{\boldsymbol{I}} - \frac{c_1}{c_0 + c_1 N}\mathbf{1}\mathbf{1}'\right),$$

which implies that

$$w_0 = \frac{\mathbf{1}^T\boldsymbol{v}}{N} - \rho\times\left(\frac{c_0}{N} + c_1\right).$$

Therefore, $\partial r_n/\partial\alpha_n = (1-1/N)\partial v_n/\partial\alpha_n$, whereas $\partial r_n/\partial\alpha_p = -N^{-1}\partial v^p/\partial\alpha_p$, $n \neq p$. Thus,

$$\frac{\partial r_n/\partial\alpha_n}{\partial r_p/\partial\alpha_n} = -(N-1) < 0, \quad n \neq p,$$

which proves the first part of the corollary.

For the second part, we note that

$$\frac{d(L_n I_n)}{d\alpha_p} = \frac{M}{\rho}\left[\Sigma^{-1}\frac{\partial(\boldsymbol{v} - w_0\mathbf{1})}{\partial\alpha_p}\right]_n,$$

which equals $\frac{M}{\rho c_0}(1-1/N)(\partial v_n/\partial\alpha_n)$ for $n = p$ and $\frac{M}{\rho c_0}(-1/N)(\partial v^p/\partial\alpha_p)$ for $n \neq p$. This implies that

$$\frac{d(L_n I_n)/d\alpha_p}{dr_n/dr_p} = \frac{M}{\rho c_0} > 0, \quad \forall n, p.$$

||

Proof of Corollary 2. Similar to the proof of Lemma 5, using the results of Proposition 1. ||

Proof of Corollary 3. We first show for an arbitrary economy, the value-weighted market portfolio will seem to be under-weighted on high-productivity firms and over-weighted on low-productivity firms. Specifically, we show that

(a) Industries with high α will seem to be under-represented in the market portfolio, *i.e.* there is an $n_0 \leq N$ such that for all $n \geq n_0$, the relative size of sector n,

$$\frac{\hat{L}^n\hat{I}^n}{\sum_j \hat{L}^j\hat{I}^j} < \frac{[\Sigma_\mu^{-1}\bar{\hat{\boldsymbol{\mu}}}]_n}{\mathbf{1}\Sigma_\mu^{-1}\bar{\hat{\boldsymbol{\mu}}}}.$$

(b) Industries with low α will seem to be over-represented in the market portfolio, *i.e.* there is an $n_0 \geq 1$ such that for all $n \leq n_0$, the relative size of sector n,

$$\frac{\hat{L}^n\hat{I}^n}{\sum_j \hat{L}^j\hat{I}^j} > \frac{[\Sigma^{-1}\bar{\hat{\boldsymbol{\mu}}}]_i}{\mathbf{1}\Sigma^{-1}\bar{\hat{\boldsymbol{\mu}}}}.$$

We define $\boldsymbol{z} = \boldsymbol{v} - w_0\mathbf{1}$. The value-weighted market portfolio is then $\boldsymbol{q} = \Sigma^{-1}\boldsymbol{z}/(\mathbf{1}^T\Sigma^{-1}\boldsymbol{z})$. The portfolio that the CAPM would predict, on the other hand, is $\boldsymbol{q}^* = \Sigma_\mu^{-1}\bar{\boldsymbol{\mu}}/(\mathbf{1}^T\Sigma_\mu^{-1}\bar{\boldsymbol{\mu}}) = \Lambda_{1-b}^{-1}\Sigma^{-1}\boldsymbol{z}/(\mathbf{1}^T\Lambda_{1-b}^{-1}\Sigma^{-1}\boldsymbol{z})$.

We therefore have

$$\frac{[\boldsymbol{q}]_i}{[\boldsymbol{q}^*]_i} = c\times(1-b_i), \quad \text{where } c = \frac{\mathbf{1}^T\Lambda_{1-b}^{-1}\Sigma^{-1}\boldsymbol{z}}{\mathbf{1}^T\Sigma^{-1}\boldsymbol{z}} > 0$$

(where the strict positivity of c is guaranteed by the strict positivity of all the elements of $\boldsymbol{L}$ since $\Sigma^{-1}\boldsymbol{z} = M^{-1}\rho\Lambda_I\boldsymbol{L}$ as showed by Proposition 2 part (vi)), which is a decreasing function of b, and therefore, as b is an increasing function of α, a decreasing function of α.

Further, as $\mathbf{1}^T\boldsymbol{q}^* = \mathbf{1}^T\boldsymbol{q} = 1$ and the portfolio weights are all non-negative, it is clear that, as long as there is dispersion of productivity $\alpha^i \neq \alpha^j$ for some i, j, there is a $\bar{\alpha} \in (\alpha_1, \alpha_N)$ such that for all industries, n, in which $\alpha_n < \bar{\alpha}$, $[\boldsymbol{q}^*]_i < [\boldsymbol{q}]_i$, and for all industries in which $\alpha_n > \bar{\alpha}$, $[\boldsymbol{q}^*]_i > [\boldsymbol{q}]_i$ ($\boldsymbol{q} \neq \boldsymbol{q}^*$) (w.l.o.g., assume that there is a i such that $[\boldsymbol{q}]_i < [\boldsymbol{q}^*]_i$. Then, there has to be a j such that $[\boldsymbol{q}]_j > [\boldsymbol{q}^*]_j$, as $\mathbf{1}^T\boldsymbol{q}^* = \mathbf{1}^T\boldsymbol{q} = 1$). Thus, low-productivity industries will indeed look "too big", whereas high-productivity industries will look "too small". This general result is true regardless of the parameters of the economy. We note that, as $c/(1-b_i) > 1$ for some i, $c < 1$.

We now turn to (i)–(ii) in the corollary. We prove the result for the case where $\Sigma = c_0\bar{\boldsymbol{I}} + c_1\mathbf{1}\mathbf{1}^T$, $c_0 > 0$, $c_1 > 0$, and $\bar{\boldsymbol{I}}$ is the identity matrix, while $\mathbf{1}$ is a vector of ones.

It is straightforward to show that the CAPM, based on the value-weighted market portfolio, predicts a vector of expected returns of

$$\boldsymbol{z}^* = d\times\Sigma\Lambda_{1-b}\Sigma^{-1}\boldsymbol{z}, \quad \text{where } d = \frac{\boldsymbol{z}^T\Lambda_{1-b}\Sigma^{-1}\boldsymbol{z}}{\boldsymbol{z}^T\Sigma^{-1}\Lambda_{1-b}\Sigma\Lambda_{1-b}\Sigma^{-1}\boldsymbol{z}},$$

whereas the true expected returns (per unit of $\tilde{\mathbf{x}}$ risk) are $\mathbf{z}$. However, using the definitions of $\boldsymbol{q}$ and $\boldsymbol{q}^*$, this is equivalent to

$$r_i \stackrel{\text{def}}{=} \frac{[\mathbf{z}^*]_i}{[\mathbf{z}]_i} = d \times c \times \frac{[\Sigma \boldsymbol{q}^*]_i}{[\Sigma \boldsymbol{q}]_i}.$$

Since $\Sigma = c_0 \bar{\boldsymbol{I}} + c_1 \mathbf{1}\mathbf{1}^T$, this implies that

$$r_i = d \times \frac{[\boldsymbol{q}]_i/(1-b_i) + c \times c_1/c_0}{[\boldsymbol{q}]_i + c_1/c_0}.$$

Now, since $c < 1$, we have $r_i < d/(1-b_i)$ $\Big($this can, *e.g.* be seen by defining $R(q) = \frac{q + c \times (1-b_i)c_1/c_0}{q + c_1/c_0}$, noting that $R(0) = c(1-b_i) < (1-b_i) < 1$, and $dR/dq = (q + c \times c_1/c_0(1-b_i))^{-1}(1-R) > 0$ iff $R > 0$, and $\lim_{q\to\infty} R(q) = 1$, so $R(q) < 1$ regardless of q, and, because $r_i = d \times R([\boldsymbol{q}]_i)/(1-b_i)$, the result follows$\Big)$.

For arbitrary i, j, such that $b_i < b_j$, we define the functions

$$Q(b) \stackrel{\text{def}}{=} [\Sigma \boldsymbol{q}]_i + \frac{[\Sigma \boldsymbol{q}]_j - \Sigma[\boldsymbol{q}]_i}{b_j - b_i}(b - b_i)$$

and

$$Z(b) \stackrel{\text{def}}{=} \frac{Q(b)/(1-b) + c \times c_1/c_0}{Q + c_1/c_0}.$$

Clearly, $Q(b) > 0$ and $Z(b) > 0$. We first show that $[\Sigma \boldsymbol{q}]_i < [\Sigma \boldsymbol{q}]_j$. From Proposition 1, b is strictly increasing in α. We have

$$\Sigma \boldsymbol{q} = (c_0 \bar{\boldsymbol{I}} + c_1 \mathbf{1}\mathbf{1}^T)\mathbf{z} = c_0 \mathbf{z} + c_1(\mathbf{1}^T \mathbf{z})\mathbf{1},$$

so

$$[\Sigma \boldsymbol{q}]_j - [\Sigma \boldsymbol{q}]_i = c_0([\mathbf{z}]_j - [\mathbf{z}]_i) > 0,$$

as $\upsilon^j > \upsilon_i$ when $\alpha^j > \alpha^i$ and $[\mathbf{z}]_j - [\mathbf{z}]_i = \upsilon^j - \upsilon_i$.

We then have $r_j - r_i = Z(b_j) - Z(b_i)$, so showing that $Z'(b) > 0$, for $b_i < b < b_j$ is enough to ensure that $r_j > r_i$. It is easy to check that

$$Z'(b) = \frac{1}{(1-b)[\boldsymbol{q}^*]_i + c_1/(c_0 \times c)}(Q'(b)(1 - c(1-b)Z(b)) + Q(b)Z(b)).$$

Moreover, since $Z(b) > 0$ and $(1-b)Z(b) < d$ (which is clearly the case since $(1-b)Z(b) = R(Q(b))$, and $d(R(Q(b))/db = R'Q' > 0$, so $(1-b)Z(b)$ reaches its maximum at b_j, at which point it is $r_j(1-b_i) < d$), if $cd < 1$, then $Z'(b) > 0$. A similar argument to the one used in showing that $c < 1$ indeed confirms that $d < 1$, so $cd < 1$, and indeed $Z'(b) > 0$, and $r_j > r_i$. Thus, r_n—the rate of CAPM-predicted expected returns to true expected returns—is an increasing function of n. A similar argument as that made when proving (a) and (b) shows that for i's such that $\alpha_i < \bar{\alpha}$, $r_i < 1$, whereas for i's such that $\alpha_i > \bar{\alpha}$, $r_i > 1$. This concludes the proof of Corollary 3. ‖

Proof of Corollary 4. Follows directly from Corollary 3. ‖

Proof of Corollary 5. Follows directly from Corollary 3. ‖

Acknowledgment. We have benefited from helpful comments by Jonathan Berk, Paul Ehling, Jacques Olivier, Dimitris Papanikolaou, Adriano Rampini, Neng Wang, Jim Wilcox, and seminar participants at Berkeley, Hautes Études Commerciales (HEC) Lausanne, HEC Paris, Institut Européen d'Administration des Affaires, Oxford Said Business School, the European Winter Finance Summit 2008, the meetings of the Western Finance Association 2008, the meetings of the Western Finance Association 2008 and the National Bureau of Economic Research spring asset pricing meeting 2008. We also thank Bruno Biais and three anonymous referees.

REFERENCES

ACHARYA, V. and BISIN, A. (2009), "Managerial Hedging, Equity Ownership and Firm Value", *Rand Journal of Economics*, **40**, 47–77.

BANSAL, R., FANG, E. and YARON, A. (2006), "Equity Capital: A Puzzle?" (Mimeo, Duke University).

BANZ, R. W. (1981), "The Relationship between Return and Market Value of Common Stocks", *Journal of Financial Economics*, **9**, 3–18.

BODIE, Z., MERTON, R. C. and SAMUELSON, W. F. (1992), "Labor Supply Flexibility and Portfolio Choice in a Life Cycle Model", *Journal of Economic Dynamics and Control*, **16**, 427–449.

BREEDEN, D. T. (1979), "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities", *Journal of Financial Economics*, **7**, 265–296.

CAMPBELL, J., CALVET, L. E. and SODINI, P. (2007), "Down or Out: Assessing the Welfare Costs of Household Investment Mistakes", *Journal of Political Economy*, **115**, 707–747.

CHAN, L., LAKONISHOK, J. and SOUGIANNIS, T. (2001), "The stock market valuation of research and development expenditure", *Journal of Finance*, **56**, 2431–2456.

DANTHINE, J. and DONALDSON, J. (2002), "Labour Relations and Asset Returns", *Review of Economic Studies*, **69**, 41–64.

FAMA, E. F. and FRENCH, K. R. (1993), "Value Versus Growth: The International Evidence", *Journal of Finance*, **53**, 1975–1999.

HEATON, J. and LUCAS, D. (1996), "Evaluating the Effects of Incomplete Markets on Risk Sharing and Asset Pricing", *Journal of Political Economy*, **104**, 443–487.

HOLMSTRÖM, B. and MILGROM, P. (1987), "Aggregation and Linearity in the Provision of Intertemporal Incentives", *Econometrica*, **55**, 303–328.

INGERSOLL, J. (1987), *Theory of Financial Decision Making* (Oxford: Rowman & Littlefield Publishers, Inc.).

JAGANNATHAN, R. and WANG, Z. (1996), "The Conditional CAPM and the Cross-section of Expected Returns", *Journal of Finance*, **51**, 3–53.

KRUSE, D. L. (1992), "Profit Sharing and Productivity: Microeconomic Evidence from the United States", *Economic Journal*, **102**, 24–26.

MALLOY, C., MOSKOWITZ, T. and VISSING-JORGENSON, A. (2005), "Job Risk and Asset Returns" (London Business School mimeo).

MAYERS, D. (1973), "Nonmarketable Assets and the Determination of Capital Asset Prices in the Absence of a Riskless Asset", *Journal of Business*, **46**, 258–267.

OU-YANG, H. (2005), "An Equilibrium Model of Asset Pricing and Moral Hazard", *Review of Financial Studies*, **18**, 1253–1303.

PALACIOS-HUERTA, I. (2003), "The Robustness of the Conditional CAPM with Human Capital", *Journal of Financial Econometrics*, **1**, 272–289.

QIN, J. (2002), "Human-Capital-Adjusted Capital Asset Pricing Model", *Japanese Economic Review*, **53**, 182–198.

ROLL, R. (1977), "A Critique of the Asset Pricing Theory's Tests Part I: On Past and Potential Testability of the Theory", *Journal of Financial Economics*, **4**, 129–176.

SANTOS, T. and VERONESI, P. (2006), "Labor Income and Predictable Stock Returns", *Review of Financial Studies*, **19**, 1–44.

TELMER, C. I. (1993), "Asset Pricing Puzzles and Incomplete Markets", *Journal of Finance*, **48**, 1803–1832.

ZAME, W. (2007), "Incentives, Contracts and Markets: A General Equilibrium Theory of Firms", *Econometrica*, **75**, 1453–1500.