Principal Agent Models with Asymmetric Information

Robert A. Miller

DSE at UCL

July 2025

A Pure Moral Hazard Model

Motivation

- In an Arrow Debreu world with a Walrasian equilibrium, whether an employee is paid the value of his marginal product less the amenity value with a certain wage or a piece rate is irrelevant.
- When does the contract form matter?
- It arises naturally in environments with *asymmetric information*.
- A *risk neutral principal* proposes a compensation plan to a *risk averse agent*.
- The agent accepts or rejects the principal's (implicit) offer.
- If he rejects the offer he receives a fixed utility from an *outside option*.
- If he accepts the offer, the agent chooses between *working* and *shirking*.
- The principal observes whether the offer is accepted, but not the agent's activity.
- After revenue is realized, the agent receives compensation according to the contract.
- The principal pockets the remainder as *profit*.

A Pure Moral Hazard Model

The agent's choices and compensation, and the principal's revenue and profits

- Denote the workplace employment decision of the agent by an indicator $l_0 \in \{0, 1\}$, where $l_0 = 1$ means the agent rejects the principal's offer.
- Denote the effort level choices by $l \in \{0, 1\}$, where work is defined by setting $l = 1$, and shirking is defined by setting $l = 0$.
- Gross revenue to the principal is denoted by x, a random variable with a probability distribution determined by the agent's effort.
- After x is revealed the both the principal and the agent at the end of the period, the agent receives compensation according to the contract.
- To reflect its potential dependence on x, we denote compensation by $w(x)$.
- The principal's profit is revenue less compensation, $x - w(x)$.

A Pure Moral Hazard Model

Marginal product of the agent

- Denote by $f(x)$ the probability density function for revenue conditional on the agent working, and let $f(x)\, g(x)$ denote the probability density function for revenue when the agent shirks.
- We assume:

$$E[xg(x)] \equiv \int xf(x)\, g(x)\, dx < \int xf(x)\, dx \equiv E[x]$$

- The inequality reflects the preference of principal for working over shirking.
- Since $f(x)$ and $f(x)\, g(x)$ are densities, $g(x)$, the ratio of the two densities, is a likelihood ratio.
- That is $g(x)$ is nonnegative for all x, bounded, and:

$$E[g(x)] \equiv \int g(x)\, f(x)\, dx = 1$$

A Pure Moral Hazard Model

Preferences of the agent

- We assume the agent is an expected utility maximizer and utility is exponential in compensation, taking the form:

$$-l_0 - l\alpha E\left[e^{-\gamma w(x)}\right] - (1-l)\,\beta E\left[e^{-\gamma w(x)} g(x)\right]$$

  where the utility of the outside option is normalized to negative one, and:
  - γ is the coefficient of absolute risk aversion.
  - α is a utility parameter with consumption equivalent $-\gamma^{-1}\log(\alpha)$ that measures the distaste from working.
  - β is defined analogously.
- We assume $\alpha > \beta$ meaning that shirking gives more utility to the agent, than working.
- A conflict of interest arises between the principal and the agent because he prefers shirking, meaning $\beta < \alpha$ yet the principal prefers working since $E[xg(x)] < E[x]$.

Solving the Pure Moral Hazard Model

Participation and incentive compatibility constraints

- To induce the agent to accept the principal's offer and engage in his preferred activity, shirking, it suffices to propose a contract that gives the agent an expected utility of at least minus one:

$$\beta E\left[e^{-\gamma w(x)} g(x)\right] \leq 1$$

- To elicit work from the agent, the principal must offer a contract that gives the agent a higher expected utility than the outside option, and a higher expected utility than shirking:

$$\alpha E\left[e^{-\gamma w(x)}\right] \leq 1$$

and:

$$\alpha E\left[e^{-\gamma w(x)}\right] \leq \beta E\left[e^{-\gamma w(x)} g(x)\right]$$

Lemma (Margiotta and Miller, 2000)

To cost minimizing contract inducing work is:

$$w^o(x) \equiv \gamma^{-1} \ln \alpha + \gamma^{-1} \ln \left[1 + \eta \left(\frac{\alpha}{\beta} \right) - \eta g(x) \right]$$

where η is the unique positive solution to the equation:

$$E\left[\frac{g(x)}{\alpha + \eta[(\alpha/\beta) - g(x)]} \right] = E\left[\frac{(\alpha/\beta)}{\alpha + \eta[(\alpha/\beta) - g(x)]} \right]$$

- The optimal contract pays the agent a:
  - fixed wage of $\gamma^{-1} \ln \beta$ to shirk.
  - certainty equivalent of $\gamma^{-1} \ln \alpha$ to work (with compensating differential $\gamma^{-1} \ln \alpha / \beta$).
  - positive risk premium of $E[w^o(x)] - \gamma^{-1} \ln \alpha$ to induce work.

Measuring the Importance of Moral Hazard

Three measures

- Recall the optimal compensation with moral hazard is $w^o(x)$ and to meet the participation constraint, shareholders must pay $\gamma^{-1}\ln\alpha$.
- Therefore the maximal amount shareholders would pay to rid the firm of the moral hazard problem is:

$$\Delta_1 \equiv E_t\left[w^o(x) - \gamma^{-1}\ln\alpha\right] = \gamma^{-1}E\left\{\ln\left[1+\eta\left(\frac{\alpha}{\beta}\right) - \eta g(x)\right]\right\}$$

- A second measure of moral hazard is the nonpecuniary benefits the manager obtains from shirking.
- This is the monetized utility loss from working versus shirking:

$$\Delta_2 \equiv \gamma^{-1}\ln\beta - \gamma^{-1}\ln\alpha = -\gamma^{-1}\ln(\alpha/\beta)$$

- Third is the gross loss a firm incurs from the manager shirking instead of working:

$$\Delta_3 \equiv E\left[x - xg(x)\right]$$

Identification

Model primitives and the data generating process

- The model is defined by:
  - $f(x)$ the probability density function of x from working
  - $g(x)$ the likelihood ratio for shirking versus working
  - α distaste for working relative to outside option
  - β distaste for shirking relative to outside option
  - γ risk-aversion parameter.
- The panel data set is $\{x_n, w_n\}_{n=1}^{N}$ where $w(x) = E[w_n | x_n]$.
- Thus $f(x)$ and $w(x)$ are identified.
- This leaves only $g(x)$ plus (α, β, γ) to identify.

Identification

What if the risk parameter is known?

- The FOC for the Lagrangian can be expressed as:

$$v(x)^{-1} = \alpha\left[1 + \eta\left(\alpha/\beta\right) - \eta g(x)\right] = \overline{v}^{-1} - \alpha\eta g(x)$$

where $v(x) = e^{-\gamma w(x)}$:

$$\lim_{x\to\infty}\left[g(x)\right] = 0 \Rightarrow \lim_{x\to\infty}\left[v(x)^{-1}\right] = \alpha\left[1 + \eta\left(\alpha/\beta\right)\right] \equiv \overline{v}^{-1}$$

- These equalities imply:

$$g(x) = \frac{\overline{v}^{-1} - v(x)^{-1}}{\alpha\eta} = \frac{\overline{v}^{-1} - v(x)^{-1}}{\overline{v}^{-1} - E\left[v(x)^{-1}\right]} \tag{1}$$

- Also since both participation and incentive compatibility constraints bind:

$$\alpha = E\left[v(x)\right]^{-1} \tag{2}$$

$$\beta = E\left[v(x)g\left(x\right)\right]^{-1} \tag{3}$$

Identification

The identified set (Gayle and Miller, 2015)

- Noting $\overline{v} \equiv e^{-\gamma \overline{w}}$ equations (1), (2) and (3) imply:

$$\alpha(\gamma) = E\left[e^{-\gamma w^o(x)}\right]^{-1}$$

$$\beta(\gamma) = \frac{1 - E\left[e^{\gamma w^o(x) - \gamma \overline{w}}\right]}{E\left[e^{-\gamma w^o(x)}\right] - e^{-\gamma \overline{w}}}$$

$$g(x, \gamma) = \frac{e^{\gamma \overline{w}} - e^{\gamma w^o(x)}}{e^{\gamma \overline{w}} - E\left[e^{\gamma w^o(x)}\right]}$$

- Finally since paying $w^o(x)$ is more profitable than paying $\gamma^{-1} \ln(\beta)$:

$$\begin{aligned} 0 &\leq E[x] - E\left[w^o(x)\right] - E[xg(x)] + \gamma^{-1} \ln(\beta) \\ &= \frac{cov\left(x, e^{\gamma w^o(x)}\right)}{e^{\gamma \overline{w}} - E\left[e^{\gamma w^o(x)}\right]} - E\left[w^o(x)\right] + \gamma^{-1} \ln\left(\frac{1 - E\left[e^{\gamma w^o(x) - \gamma \overline{w}}\right]}{E\left[e^{-\gamma w^o(x)}\right] - e^{-\gamma \overline{w}}}\right) \end{aligned}$$

- The set of γ satisfying this inequality is *sharp* and *tight*, and the model is rejected if this set is empty.

A Dynamic Extension to the Static Model

.

- Adding *simple dynamics* to this model further restricts the set of *observationally equivalent* parameterizations.
- In a multiperiod model where the agent can borrow and save:
  - *interest rate adjustments* affect the value of (smoothing) an extra dollar
  - shifting the incentive compatibility and participation constraints (Gayle and Miller, 2009)
- Accordingly suppose that each period t:
  - the agent chooses his consumption c_t.
  - the principal announces a compensation function $w_t(x_{t+1})$.
  - the agent chooses $l_{0t} \in \{0, 1\}$ (participation) and $l_t \in \{0, 1\}$ (effort).
  - Output x_{t+1} occurs and he is paid
- For some discount factor $\delta \in (0, 1)$ his lifetime utility is:

$$-\sum_{t=0}^{\infty} \delta^t \exp\left(-\gamma c_t\right)\left[l_{0t} + l_t \alpha + \left(1 - l_t\right)\beta\right]$$

where the preference parameters (α, β, γ) and the production parameters $(f(x), g(x))$ have the same interpretation as above.

A Dynamic Extension to the Static Model

Modifying the participation and incentive compatibility constraints

- Similar to the static model define:

$$v_t(x) \equiv \exp\left(-\gamma w_t(x) / b_{t+1}\right)$$

  where b_t denote the bond price, and assume b_{t+1} is known at period t
- One can show the participation and incentive-compatibility constraints also follow their static model analogues:

$$\alpha^{-1/(b_t - 1)} \geq E\left[v_t(x)\right] \tag{4}$$

$$0 \geq E\left[\left(g(x) - (\alpha/\beta)^{1/(b_t - 1)}\right) v_t(x)\right]. \tag{5}$$

- The principal chooses v_t for each x to maximize:

$$\int \ln\left[v_t(x)\right] f(x) \mathrm{d}x$$

  subject to (4) and (5).
- Changes in b_t tilt the constraints through the effect on $v_t(x)$.
- Since b_t is exogenous, it is a instrument facilitating identification.

A Dynamic Extension to the Static Model

Short term contracts are optimal (Proposition 5, Margiotta and Miller, 2000)

Lemma

The optimal long-term contract is implemented by replicating optimal short-term contracts, where the agent retires for sure in period t or $t+1$, choosing (l_{t0}, l_{t1}, l_{t2}) to maximize:

$$-l_{0t} - E_t \left\{ \left[l_t \alpha^{1/(b_t-1)} + (1-l_t)\,\beta^{1/(b_t-1)} g_t(x) \right] \exp\left(-\frac{\gamma w_t(x)}{b_{t+1}} \right) \right\}$$

- Comparing the principal's problem in this dynamic setting to its static analogue, the only differences are that:
  - $\gamma \big/ b_{t+1}$ replaces γ, the risk aversion parameter. *Idiosyncratic wealth shocks are optimally smoothed over the agent's lifetime.*
  - $\alpha^{1/(b_t-1)}$ replaces α and $\alpha^{1/(b_t-1)}$ replaces α. *The consumption equivalent of $\alpha^{1/(b_t-1)}$ is $[(b_t-1)\gamma]^{-1} \ln \alpha$, augmenting (reducing) wealth when $\alpha \lesseqgtr 1$.*
- This result builds on Malcomson and Spinnewyn (1988), and Fudenberg, Holmstrom and Milgrom (1990).
- Intuitively, no information about shirking in t arrives after $t+1$, so postponing rewards or penalties beyond one period is pointless.

A Fully Parametric Specification

Truncated Normal distribution and Absolute Risk Aversion (CARA)

- Assume x is distributed truncated normal with lower truncation point ψ (representing bankruptcy or limited liability) with mean μ_w (μ_s) and variance σ^2 for parent normal if agent works (shirks):

$$
\begin{aligned}
f(x) &= \frac{1}{\sigma_w \sqrt{2\pi}} \Phi\left(\frac{\mu_w - \psi}{\sigma}\right)^{-1} \exp\left[\frac{-(x-\mu_w)^2}{2\sigma^2}\right] \\
\ln g(x) &= \ln \Phi\left[(\mu_s - \psi)/\sigma\right] - \ln \Phi\left[(\mu_w - \psi)/\sigma\right] \\
&\quad + \frac{\mu_w^2 - \mu_s^2}{2\sigma^2} + \frac{(\mu_s - \mu_w)}{\sigma^2} x
\end{aligned}
$$

- Thus the model is parameterized by $(\psi, \mu_w, \sigma, \mu_s, \gamma, \alpha, \beta)$.
- Suppose there are N observations on $(\widetilde{w}_n, x_n)$ where:

$$\widetilde{w}_n \equiv w_n + \epsilon_n \text{ and } E\left[\epsilon_n \,|x_n\right] = 0.$$

- Margiotta and Miller (2000) estimate:
  1. ψ with $\widehat{\psi} \equiv \min\{x_1, \ldots, x_N\}$. (Note $\widehat{\psi}$ converges to ψ at rate faster than $\sqrt{N}$ but is sensitive to measurement error.)
  2. (μ_w, σ) with LIML by forming likelihood for $f(x)$ with $\{x_1, \ldots, x_N\}$ under the assumption that $\widehat{\psi} = \psi$. (No first stage correction is necessary.)
  3. $(\mu_s, \gamma, \alpha, \beta)$ with NLS based on

     $$\widetilde{w}_n = \gamma^{-1} \ln \alpha + \gamma^{-1} \ln \left[1 + \eta \left(\frac{\alpha}{\beta} \right) - \eta g(x) \right] + \epsilon_n$$

     using an inner loop at each iteration to solve for η as a mapping of (α, β, μ_s) given $(\widehat{\psi}, \widehat{\mu}_w, \widehat{\sigma})$.
  4. Correct the standard errors for $(\mu_s, \gamma, \alpha, \beta)$ in the third step induced by $(\widehat{\mu}_w, \widehat{\sigma})$ obtained from the second step.

A Fully Parametric Specification

Estimating the importance of moral hazard (Table 8, Margiotta and Miller 2000)

- We used the Masson-Antle-Smith (MAS) data set (37 firms in aerospace, electronics, chemicals from 1944 - 1977).
- The annual cost of moral hazard pales in comparison to losses shareholders would make if managers were paid a fixed wage.

Measure	Industry	Executive	Cost
Δ_1	Aerospace	CEO	186,689
		Non-CEO	2,370
	Chemicals	CEO	232,966
		Non-CEO	2,680
	Electronics	CEO	173,643
		Non-CEO	2,327
Δ_2		CEO	259,181
		Non-CEO	3,272
Δ_3	Aerospace		263,283,500
	Chemicals		85,355,000
	Electronics		104,222,000

50 Years of Managerial Compensation

Changes in managerial compensation (Table 3, Gayle and Miller, 2009)

- We compare MAS data with data from:
  - S&P 500 COMPUSTAT CRSP (2,610 firms 1995 -2004, 2000 $US)
  - A subset formed from those firms in the three MAS sectors.

Rank	Sector	Old	New restricted	New all
All	All	528 (1,243)	4,121 (19,283)	2,319 (12,121)
CEO	All	729 (1,472)	6,109 (24,250)	5,320 (19,369)
Non-CEO	All	400 (1,026)	2,256 (12,729)	1,562 (9,303)
All	Aerospace	744 (1,140)	6,407 (20,689)	
CEO	Aerospace	950 (1,292)	11,664 (19,416)	
Non-CEO	Aerospace	624 (695)	1,997 (18,563)	
All	Chemicals	543 (1,348)	2,802 (9,555)	
CEO	Chemicals	718 (1,527)	3,673 (7,072)	
Non-CEO	Chemicals	401 (241)	477 (23,390)	
All	Electronics	370 (1,057)	4,501 (22,118)	
CEO	Electronics	457 (1,407)	5,325 (24,576)	
Non-CEO	Electronics	108 (61)	1,635 (18,810)	

50 Years of Managerial Compensation

Changes in components of managerial compensation (Table 4, Gayle and Miller, 2009)

Variable	Rank	Old	New restricted	New all
Salary and bonus	All	219	838	667
		(114)	(1,066)	(905)
	CEO	261	1,037	1,127
		(115)	(1,365)	(1,282)
	Non-CEO	179	640	552
		(97)	(576)	(738)
Value of options granted	All	79	2,401	903
		(338)	(13,225)	(3,753)
	CEO	111	3,402	1,782
		(439)	(18,172)	(7,169)
	Non-CEO	51	1,401	681
		(198)	(4,237)	(2,106)
Value of restricted stock granted	All	11	187	152
		(95)	(1,633)	(936)
	CEO	8	242	298
		(72)	(2,021)	(1,464)
	Non-CEO	13	133	115
		(112)	(1,118)	(743)
Change in wealth from options held	All	5	785	281
		(134)	(14,636)	(8,710)
	CEO	7	1,667	1,474
		(167)	(17,078)	(13,567)
	Non-CEO	3	−76	−18
		(94)	(11,706)	(6,939)
Change in wealth from stock held	All	−3	−40	125
		(439)	(5,681)	(4,350)
	CEO	0.434	−14	264
		(479)	(6,712)	(6,791)
	Non-CEO	−7	−64	90
		(398)	(4,496)	(3,473)

50 Years of Managerial Compensation

Changes in sample composition of firms (Table 2, Gayle and Miller, 2009)

Variable	Sector	Old	New restricted	New all
Sales	All	1,243 (2,250)	3,028 (6,830)	4,168 (109,000)
	Aerospace	1,886 (3,236)	11,500 (14,900)	
	Chemicals	1,246 (2,018)	2,252 (2,091)	
	Electronics	319 (536)	2,469 (6,223)	
Value of equity	All	589 (1,034)	1,273 (2,863)	1,868 (4,648)
	Aerospace	391 (680)	3,132 (3,826)	
	Chemicals	677 (1,107)	800 (869)	
	Electronics	159 (365)	1,283 (3,096)	
Number of firms	All	37	151	1,517
	Aerospace	5	11	
	Chemicals	25	40	
	Electronics	7	100	
Number of employees	All	27,370 (28,850)	12,208 (26,676)	18,341 (46,960)
	Aerospace	49,920 (34,335)	58,139 (69,452)	
	Chemicals	23,537 (25,268)	8,351 (9,323)	
	Electronics	10,485 (7,664)	9,195 (18,266)	
Total assets	All	525 (924)	3,035 (6,550)	9,926 (40,300)
	Aerospace	726 (130)	10,600 (12,900)	
	Chemicals	548 (851)	2,385 (2,380)	
	Electronics	146 (233)	2,551 (6,311)	
Observations	All	1,797	3,260	82,578
	Aerospace	355	233	
	Chemicals	1,092	935	
	Electronics	252	2,092	

50 Years of Managerial Compensation

What were the driving forces behind these changes?

- If managers in the COMPUSTAT population ran firms the same size as managers in MAS, their compensation would have increased by a factor of 2.3, the increase in national income per capita.
- After adjusting for the general increase in living standards over these years, the model attributes:
  - Hardly any of the increased managerial compensation to changes in $\gamma^{-1} \ln \alpha$, or the certainty equivalent wage
  - practically all the increase to changes the risk premium Δ_1
- The factors driving the change in Δ_1 were:
  - not risk preferences: managers in the MAS (COMPUSTAT) population were willing to \$240,670 (\$248,620) to avoid a gamble of winning or losing \$1 million.
  - not changes in $f(x)$: the biggest change in Δ_1 in aerospace where the abnormal returns became less dispersed *(reducing the premium)*.
  - the sharp increase in α / β mainly due to increased firm assets *(increasing the utility from shirking)*.

- The last part of this lecture extends the analysis by developing a generalized *Roy model*:
  - in a *dynamic* setting.
  - where there is one *principal* and *multiple agents* in each firm.
  - Each agent has several *employment choices*,
  - and accumulates *human capital*.
- We apply this model to managerial compensation to explain **why executives in large firms are paid more than those in small firms**.

Data

Sources and summary statistics (Gayle, Golan and Miller, 2012)

- Data taken from ExecuComp for the S&P 1500 and COMPUSTAT were matched with data from Who's Who for the years 1992-2006:
  - 16,300 executives (from 30,614) in 2100 firms (from 2818) yielding 59,066.
- Information on executives includes:
  - compensation, title, including interlock status, age, gender, education, annual transitions by title and firm.
- Information on firms include:
  - annual financial return, size by total assets (*large, medium, small*) and sector (*primary, service, consumer*).
- Summarizing some aggregates:
  1. The executive exit rate is between 12% and 18% per year.
  2. Turnover is about 2% to 3% per year.
  3. Executives average between 51 and 54 years old.
  4. On average executives have about 13 to 14 years firm tenure.
  5. They average about 17 years executive experience.
  6. About 80% graduated from college and about 20% have an MBA.
  7. Total compensation averages between \$1.5 and \$4.5 million.
  8. Compensation increases with firm size.

Compensation, education and tenure by firm size (Figures 1 and 2, GGM 2015, pages 2302-2303)

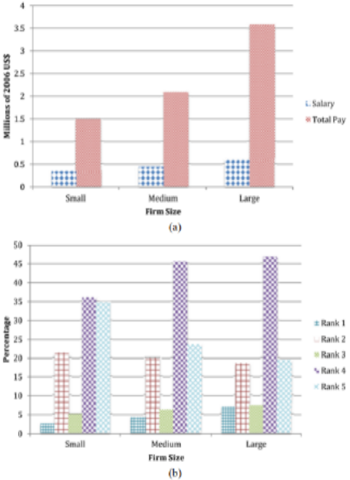

FIGURE 1.—Pay and hierarchy by firm size. (a) Firm size pay premium, (b) hierarchy by firm size.

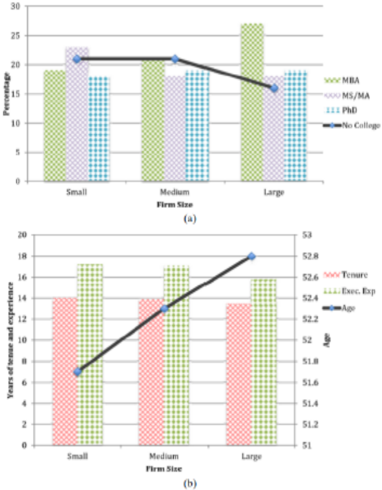

FIGURE 2.—Education and experience by firm size. (a) Education and firm size, (b) experience and firm size.

Data

What explains firm-size pay premium in the market for top executives?

- There are two fundamental factors that might be playing a role:

1. **Human Capital**:
    1. Executives in large firms are older, more educated, but have less executive experience and less tenure than those in smaller firms; presumably human capital of the kind described by Mincer (1974) is playing a role.
    2. Working as executives in more firms increases an executive compensation at higher ranks in the hierarchy. This is a form of productivity enhancing on-the-job experience.
2. **Moral Hazard**:
    1. Top executives are paid a significant portion of their total compensation in stock and options.
    2. The composition of firm denominated securities varies substantially across ranks and executives at different points in their lifecycle.

Model

Choices

- Each period while employed the executive chooses:
  - consumption $c_t \in \mathcal{R}$,
  - a job $d_{jkt} \in \{0, 1\}$,
  - and (if she does not retire) effort $l_t \in \{0, 1\}$,
- where:
  - $j \in \{1, 2..., J\}$ denotes firm size, industrial sector, new firm or not,
  - $k \in \{1, \ldots, K\}$ denotes job rank,
  - $l_t = 1$ designates working and $l_t = 0$ means shirking,
  - she retires by setting $d_{0t} = 1$,
  - and subject to the restrictions that for all $d_t \equiv (d_{0t}, d_{11t}, \ldots, d_{JKt})$:

$$d_{0t} + \sum_{j=1}^{J} \sum_{k=1}^{K} d_{jkt} = 1.$$

Model

Human capital

- The executive adds to:
  1. h_{1t}, her internal human capital, by working at the same firm,
  2. h_{2t}, her general human capital, by not quitting and working,
  3. h_{3t}, her external capital, by switching firms and working.
- Denote by:
  - h_0 is a fixed set of individual characteristics including gender and education
  - $h_t \equiv (t, h_0, h_{1t}, h_{2t}, h_{3t})$
  - $\underline{H}(h_t)$ as human capital in $t+1$ from shirking in t
  - $\overline{H}(h_t)$ as human capital from working.
- Let $d_t^* = 0$ ($d_t^* = 1$) denote employment at the same (a new) firm and assume:

$$\underline{H}(h_t) = \begin{pmatrix} t+1 \\ h_0 \\ d_t^* h_{1t} \\ h_{2t} \\ h_{3t} \end{pmatrix} \quad \overline{H}(h_t) = \begin{pmatrix} t+1 \\ h_0 \\ d_t^*(h_{1t}+1) + (1-d_t^*) \\ h_{2t} + (1-d_{0t}) \\ h_{3t} + d_t^* \end{pmatrix}$$

Model

Preferences and budget constraint

- Lifetime utility of an executive is parameterized as:

$$-\sum_{t=1}^{\infty}\sum_{j=0}^{J}\sum_{k=1}^{K}\delta^{t}e^{-\gamma c_t-\varepsilon_{jkt}}d_{jkt}\left[\alpha_{jk}\left(h_t\right)l_t+\beta_{jk}\left(h_t\right)\left(1-l_t\right)\right]$$

  where:

  - we abbreviate by setting $d_{0kt}\equiv d_{0t}$ for all k.
  - c_t is consumption at time t.
  - δ is the subjective discount factor.
  - γ denotes the coefficient of absolute risk aversion.
  - $\alpha_{jk}\left(h_t\right)$ is a preference parameter for working.
  - $\beta_{jk}\left(h_t\right)$ is a preference parameter for shirking.
  - Utility depends on h_t and:

$$\alpha_{jk}\left(h_t\right)>\beta_{jk}\left(h_t\right)>0$$

- An *iid* firm-job privately observed taste shock ε_{jkt} also affects utility.
- There are complete markets for all publicly disclosed events, but no borrowing against future executive compensation.

- Firm production is defined as:

$$\sum_{k=1}^{K} F_{jkt(\tau)}\left(h_{t(\tau)}\right) + e_{j\tau}\left(\pi_{\tau+1} - 1\right) + e_{j\tau}\pi_{j,\tau+1}$$

where for expositional ease, each executive holds a distinct position and:

  - $t(\tau)$ is the age of executive at calendar time τ
  - $h_{t(\tau)} = h_{t(\tau)}$ denotes human capital of the executive aged at time τ
  - $F_{jk,t(\tau)}\left(h_{t(\tau)}\right)$ denote the individual contribution of k to the firm
  - $e_{j\tau}$ denotes the value of firm j at the beginning of calendar time τ
  - $\pi_{\tau+1}$ denotes the gross returns to the market portfolio
  - $\pi_{j,\tau+1}$, denotes abnormal return to the firm before executive compensation.

- We assume the probability density for $\pi_{j,\tau+1}$ is:
  - $f_j(\pi_{j,\tau+1})$ when all K executives work
  - $f_j(\pi_{j,\tau+1})g_{jk}(\pi_{j,\tau+1} \,|h_t\,)$ when all executives but k work.

- The gross expected return to a firm is higher if everybody works:

$$\int \pi f_j\left(\pi\right) d\pi > \int \pi f_j\left(\pi\right) g_{jk}(\pi \,|h_t\,) d\pi$$

Model

Timing, information, and overview of perfect equilibrium

1. Each executive knows her h_t and privately chooses consumption c_t.
2. Shareholders have beliefs about h_t, namely h'_t.
3. The executive privately observes ε_{jkt} and selects a firm and position.
4. Executives in each firm simultaneously submit compensation proposals, denoted by $w_{jkt+1}(h', \pi)$, to the shareholder board.
5. If any proposal is off the equilibrium path, shareholders believe the worst and reject all the submissions.
6. This rejection is observed by all firms.
7. If their demands are not approved, the executives in the firm retire.
8. If approved, the executives privately choose l_t.
9. h_t is updated with $\underline{H}(h_t)$ or $\overline{H}(h_t)$.
10. The equilibrium optimal contract induces executives to work.

Firm and Job Choices

Indexing the value of human capital

- Recursively define $B_t\left(h, h'\right)$ an index of human capital by:

$$\begin{aligned} B_t\left(h, h'\right) &= p_{0t}\left(h, h'\right) E\left[\exp\left(-\varepsilon_{0t}^* / b_t\right)\right] \\ &+ \sum_{j=1}^{J} \sum_{k=1}^{K} p_{jkt}\left(h, h'\right) E\left[\exp\left(-\varepsilon_{jkt}^* / b_t\right)\right] V_{jkt}\left(h, h', b_t\right) \end{aligned}$$

where $V_{jkt}\left(h, h', b_t\right) \equiv$

$$\min \left\{ \begin{array}{c} \left[\alpha_{jkt}\left(h\right)\right]^{\frac{1}{b_t}} \left\{B_{t+1}\left[\overline{H}\left(h\right), \overline{H}\left(h'\right)\right] E_t\left[v'_{jk,t+1}\right]\right\}^{1-\frac{1}{b_t}}, \\ \left[\beta_{jkt}\left(h\right)\right]^{\frac{1}{b_t}} \left\{B_{t+1}\left[\underline{H}\left(h\right), \overline{H}\left(h'\right)\right] E_t\left[v'_{jk,t+1}\right]\right\}^{1-\frac{1}{b_t}} \end{array} \right\}$$

where:

- b_t is the bond price at t.
- $v'_{jk,t+1} = e^{-\gamma w_{jk,t+1}(h',\pi)/b_{t+1}}$ is the annuitized util value of compensation.
- ε_{jkt}^* is the value of the private disturbance ε_{jkt} conditional on $d_{jkt} = 1$.
- $p_{jkt}(h, h')$ is the CCP for choosing rank k in firm j, period t.

- Lower values of $B_t\left(h, h'\right)$ are associated with higher values of human capital.

Firm and Job Choices

Optimization (Theorems 4.2 and 5.1 of GGM 2015)

- The value function is derived in two steps, solving for:
  1. optimal consumption given any career path
  2. the optimal career path.
- Along the equilibrium path $h = h'$ and we define:
  - $A_{t+1}(h) = B_{t+1}\left[\overline{H}(h), \overline{H}(h)\right]$.
  - $p_t(h) = p_t(h, h)$
- In the second step (along the equilibrium path) jobs are chosen to maximize:

$$\sum_{j=0}^{J}\sum_{k=0}^{K} d_{jkt}\left\{\varepsilon_{jkt} - \ln\alpha_{jkt}(h) - (b_t - 1)\left(\ln A_{t+1}(h) + \ln E_t[v_{jk,t+1}]\right)\right\} \tag{6}$$

- Executives trade off (j, k, l), jobs and effort, based on three dimensions:
  1. nonpecuniary benefit, both idiosyncratic, ε_{jkt}, and systematic, $\alpha_{jkt}(h)$;
  2. human-capital accumulation, $A_{t+1}(h)$;
  3. expected utility from compensation, $E_t[v_{jk,t+1}]$.

Cost Minimization

Participation constraint

- By the inversion theorem (Hotz and Miller 1993) there exists $q(p)$ to R^{JK} such that:

$$q_{jk}[p_t(h)] = \ln\left[\alpha_{jkt}(h)\right] + (b_t - 1)\left\{\ln A_{t+1}(h) + \ln E_t[v_{jk,t+1}]\right\} \quad (7)$$

where $q_{jk}[p_t(h)] \equiv \varepsilon'_{jkt} - \varepsilon'_{0t}$, for all shock pairs $\left(\varepsilon'_{0t}, \varepsilon'_{jkt}\right)$ making the executive indifferent between retiring and (j, k).

- Define $w^*_{jk,t+1}(h)$ as the certainty equivalent wage to a executive indifferent between $(j.k)$ and retirement given CCPs $p_t(h)$:

$$q_{jk}[p_t(h)] = \ln\alpha_{jkt}(h) + (b_t - 1)\left\{\begin{array}{l}\ln A_{t+1}(h) \\ + \ln E_t[\exp\left(-\gamma w^*_{jk,t+1}(h) / b_{t+1}\right)]\end{array}\right\}$$

- Solving for $w^*_{jk,t+1}(h)$ gives the participation constraint:

$$w^*_{jk,t+1}(h) = \frac{b_t}{\gamma}\left\{\frac{1}{(b_t - 1)}\ln\alpha_{jkt}(h) + \ln A_{t+1}(h) - \frac{1}{(b_t - 1)} q_{jk}[p_t(h)]\right\}$$

Cost Minimization

Incentive compatibility constraint

- In this model the firm can deter shirking in a one-period contract by offering a compensation schedule that satisfies the incentive-compatibility constraint:

$$\left[\frac{\alpha_{jkt}(h)}{\beta_{jkt}(h)}\right]^{1/(b_t-1)} \leq \frac{E_t\left[v_{jk,t+1}g_{jkt}(\pi \mid h)\right] B_{t+1}\left[\overline{H}(h), \overline{H}(h')\right]}{E_t\left[v_{jk,t+1}\right] B_{t+1}\left[\underline{H}(h), \overline{H}(h')\right]}. \quad (8)$$

- A fixed wage is optimal if career concerns predominate:

$$\begin{aligned} & \ln \alpha_{jkt}(h) + (b_t - 1) \ln B_{t+1}\left[\overline{H}(h), \overline{H}(h')\right] \\ \leq \quad & \ln \beta_{jkt}(h) + (b_t - 1) \ln B_{t+1}\left[\underline{H}(h), \overline{H}(h')\right] \end{aligned} \quad (9)$$

- However if (as a counterfactual) $\underline{H}(h) = \overline{H}(h)$ then a fixed wage contract:
  - simplifies the right side of the (8) to one
  - violating (8) because $\alpha_{jkt}(h) > \beta_{jkt}(h)$
  - proving a constant wage guarantees shirking if $\underline{H}(h) = \overline{H}(h)$.

Cost Minimization

Optimal Contract (Theorem 4.3 of GGM 2015)

- The cost minimizing contract is:

$$\begin{aligned} w_{jk,t+1}(h,\pi) &= w^*_{jk,t+1}(h) + r_{jk,t+1}(h,\pi) \\ &\equiv \Delta^{\alpha}_{jkt}(h) + \Delta^{A}_{jkt}(h) + \Delta^{q}_{jkt}(h) + r_{jk,t+1}(h,\pi) \end{aligned}$$

1. $\Delta^{\alpha}_{jkt}(h) \equiv \gamma^{-1}(b_t - 1)^{-1} b_{t+1} \ln \alpha_{jkt}(h)$ is the systematic component of non-pecuniary utility of (j,k)
2. $\Delta^{A}_{jkt}(h) \equiv \gamma^{-1} b_{t+1} \ln \{A_{t+1}(h)\}$ is the investment value of (j,k).
3. $\Delta^{q}_{jkt}(h) \equiv \gamma^{-1}(b_t - 1)^{-1} b_{t+1} q_{jk}[p_t(h)]$ are the idiosyncratic values making executive in fractal $p_{jkt}(h)$ indifferent between (j,k) and retirement.
4. $\Delta^{r}_{jkt}(h)$ is the risk premium defined as:

$$\Delta^{r}_{jkt}(h) \equiv E\left[r_{jk,t+1}(h,\pi)\right] = \frac{b_{\tau+1}}{\gamma} E\left[\ln\left\{1 - \eta g_{jkt}(\pi\,|h) + \eta\left[\frac{\alpha_{jkt}(h)}{\beta_{jkt}(h)}\right]^{1/(b_\tau - 1)}\right\}\right.$$

with η the unique positive root to:

$$\int\left\{\eta^{-1} + \left[\frac{\alpha_{jkt}(h)}{\beta_{jkt}(h)}\right]^{1/(b_t-1)} - g_{jkt}(\pi\,|h)\right\}^{-1} f_j(\pi)\, d\pi = 1$$

Identification and Estimation

Compensating differentials and risk aversion

- Note that (6) is a dynamic discrete choice problem.
- Assuming $T1EV$ and denoting by $\Gamma[\cdot]$ the complete gamma function:

$$B_t\left(h, h'\right) = p_{0t}\left(h, h'\right)\Gamma\left[1 + \frac{1}{b_{t+1}}\right]$$

- Equation (7) then implies the participation constraint can be expressed as:

$$\begin{aligned} \ln\left(\frac{p_{jkt}(h)}{p_{0t}(h)}\right) \quad = \quad & -\ln\alpha_{jkt}(h) - \frac{b_t - 1}{b_{t+1}}\ln p_{0,t+1}\left[\overline{H}(h_t)\right] \\ & -(b_t - 1)\ln\Gamma\left[1 + \frac{1}{b_{t+1}}\right] - (b_t - 1)\ln E_t[\upsilon_{jk,t+1}] \end{aligned} \tag{10}$$

- Appealing to Arcidiacono and Miller (2020), $\alpha_{jkt}(h)$ and ρ are identified up the distribution of ε_t.
- Intuitively both are identified off from the different characteristics their job choices, inducing executives to reveal their attitude towards risk, the value they place on nonpecuniary features of the job, and their investment value.
- Sample analogs were constructed for the CCPs, compensation schedule, and conditional and unconditional densities of the abnormal return.
- A GMM estimator can be constructed from moment conditions using (10).

Estimates from the Structural Model

Figure 3 from GGM 2015, page 2345

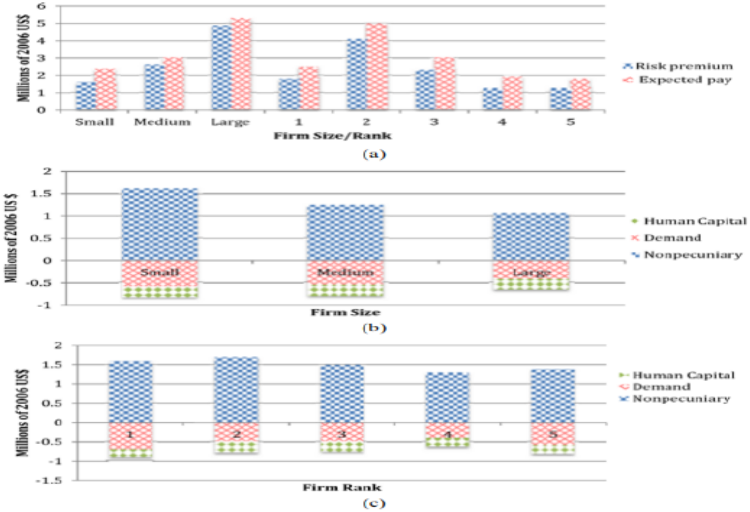

FIGURE 3.—Rank and firm-size pay decomposition. (a) Risk premium, (b) decomposition of certainty-equivalent pay, (c) decomposition of certainty-equivalent pay. *Note*: The certainty equivalent is the sum of human capital, demand, and nonpecuniary compensating differentials.

Estimates from the Structural Model

Factors explaining the firm-size executive pay premium

1. Large firms employ more talented executives.
2. There is no support for the hypothesis that executives prefer working in small firms. (*They are willing to work in a large firm for less pay.*)
3. There is no firm-size premium for human capital. (*Education and experience gained from different firms are individually significant, but collectively the firm-size pay differentials net out.*)
4. 80% of the firm-size total-compensation gap comes from the risk premium:
    - Signal quality about effort is unambiguously poorer in larger firms, and this fully explains the larger risk premium.
    - Larger firms having more supervisory positions and accountability is more difficult.
5. The remaining 20% comes from demand. Large firms pay a premium to meet demand because their bigger resource base amplifies the marginal productivity of their executives.