Working Paper 9301

SHARING WITH A RISK-NEUTRAL AGENT

by Joseph G. Haubrich

March 1993

Abstract

In the standard solution to the principal-agent problem, a risk-neutral agent bears all the risk. This paper shows that, in fact, multiple solutions exist, and often the risk-neutral agent is not the sole bearer of risk. Furthermore, as risk aversion approaches zero, the unique risk-averse solution converges to the risk-neutral solution wherein the agent bears the *least* amount of risk. Even a small degree of risk aversion can lead to agents' bearing significantly less risk than the simple solution suggests.

I. Introduction

In the classical principal-agent problem, a risk-neutral agent bears all the risk. This solution, while correct, is misleading. Other solutions exist wherein the agent does not bear all the risk, and these have some claim to being more "natural," since they are the limits of solutions to the risk-averse case.

Specifically, in the Grossman-Hart (1983; hereafter referred to as GH) principal-agent problem with a finite number of actions and states, many optimal sharing rules exist; in only one does the agent bear all the risk.[1] With a large enough stake in the project, the agent will not shirk -- and with a finite number of states and actions, this stake need not be 100 percent.

Once agents have some risk aversion, the principal-agent problem has a unique solution. For the two-state case, the limits can be computed as risk aversion approaches zero. The risk-averse solutions do not converge to the classic risk-neutral solution, however, but to the solution with the lowest risk for the agent. Less risk makes a risk-averse agent happier, so he demands a lower risk premium, in turn making the principal happier. But exceptions occur. There are knife-edge cases in which the optimal action discretely shifts with an infinitesimal increase in risk aversion. In this case, the sharing rule, and thus the risk borne by the agent, differ substantially when the principal wants to induce distinctly different actions.

By increasing the number of actions, these results reduce to the standard continuous-action principal-agent models (see Holmstrom [1979]). Under reasonable conditions, the set of risk-neutral solutions shrinks to one.

This should introduce a note of caution to applications of the principal-agent model. The simple risk-neutral solution is not a good approximation of the optimal contract, even for arbitrarily low risk aversion. It can be seriously misleading to compare actual contracts in which risk aversion is important, say in executive compensation, with the predictions of the risk-neutral principal-agent model . Fortunately, the GH model used

1

here can also deliver quantitative predictions, allowing more direct comparisons (Haubrich [1991]).

II. Other Sharing Rules

A. The Model

First, let us quickly review the assumptions, notation, and approach of the GH model. For concreteness, assume the principal owns a firm, but that she delegates its management to the agent. There are a finite number of outcomes (gross profit states) $q_1 < q_2 < ... < q_n$. The principal, who is risk neutral, only cares about the firm's expected net profit, defined as gross profit minus any payment to the manager.

In managing the firm, the agent takes an action, often thought of as effort, that the principal cannot observe. The principal does observe the outcome, however, and, like the agent, knows how different actions determine the probability of the outcome states. Both know $\pi_i(a)$, the probability of outcome q_1 given action a. This probabilistic setting means the agent might work hard but still have little output to show for it. In choosing an action, the agent does not know the ultimate result. Conversely, in seeing the outcome, the principal cannot deduce the agent's action.

Actions belong to the finite set $A=\{a_1, a_2, a_3 ... a_m\}$, making the expected benefit to the principal from an action equal to $B(a)=\sum_{i=1}^{n} \pi_i(a)q_i$. To avoid the Mirrlees (1976) problem of increasingly bigger penalties imposed with progressively smaller probabilities, we assume that $\pi_i(a)$ is strictly greater than zero for all states and actions.

The agent likes income but dislikes effort. His utility function $U(a,I)$ depends positively on his income from the principal, I, and negatively on his action, a. GH find it useful to place the following restrictions on $U(a,I)$:

Assumption A1: $U(a,I)$ has the form $G(a) + K(a)V(I)$, where $V(I)$ is a real-valued, continuous, strictly increasing, and concave function with domain

2

$[\underline{I},\infty]$ and $\lim\limits_{I\to\underline{I}} V(I) = -\infty$. G and K are real-valued, continuous functions defined on A, with K strictly positive. In addition, for all a_1, a_2 in A and I, J in $(\underline{I},\infty)$, $G(a_1) + K(a_1)V(I) \geq G(a_2) + K(a_2)V(I)$ implies $G(a_1) + K(a_1)V(J) \geq G(a_2) + K(a_2)V(J)$.

The agent has a reservation utility $\overline{U}$, the expected utility he can achieve working elsewhere. Sometimes, we consider this as derived from an outside income $\overline{I}$, so that $\overline{U}$ = $V(\overline{I})$. If the principal does not offer him a contract worth at least $\overline{U}$, the agent takes another job. To make the model at all interesting, some income level should induce the agent to work. GH formalize this as

Assumption A2: $[\overline{U}-G(a)]/K(a) \leq V(\infty)$ for all a in A.

As an example of when this assumption does not hold, consider negative exponential utility $-e^{-k(I-a)}$ and a $\overline{U}$ of $+5$. In this case, even infinite income could not make the agent work.

If the principal could observe actions, it would be straightforward to determine what she pays the agent for a particular one. Call this the first-best cost, or $C_{FB}(a)$:

$$U[a,C_{FB}(a)] = \overline{U}, \text{ or } C_{FB}(a) = h\{[\overline{U}-G(a)]/K(a)\},$$

where $h = V^{-1}$.

As GH put it (p. 11), "$C_{FB}(a)$ is simply the agent's reservation price for picking action a." Given this cost, the first-best optimal action maximizes the principal's net benefit, $B(a)-C_{FB}(a)$.

Of course, the principal cannot observe the agent's actions, nor can she directly base pay on effort. Instead, she chooses an incentive scheme $I=\{I_1,I_2,...I_n\}$ wherein payment I_i depends on the observed final state q_i. Given this, the agent will choose the

3

action that maximizes his expected utility. Knowing how the agent will react, the principal now can break her problem into two parts. For each action, she calculates the least costly incentive scheme that induces the agent to choose that course. This gives her the expected cost of motivating the agent to perform a particular action a, $C(a) = \sum_{i=1}^{n} \pi_i(a)I_i$.

She then chooses the action with the highest net benefit; that is, the one that maximizes $B(a)-C(a)$.

B. Multiple Solutions

The possibility of multiple solutions arises from looking at the mathematics of the agent's problem. With risk neutrality, the concave programming problem with a unique solution becomes a linear programming problem with multiple solutions. When neither the principal nor the agent cares about risk, risk enters only as it reflects the share held for incentive purposes. When the principal is not indifferent between the two most desirable actions, multiple equilibria can result. The larger the gap between the actions, the more risk the agent can bear. With a risk-averse agent, the principal would minimize the agent's risk subject to meeting the incentive constraints. For a risk-neutral agent, only the incentives matter, and any risk configuration consistent with them works.

The traditional solution assigns all the risk to the agent:

$$I_i = q_i - [B(a^*) - C_{FB}(a^*)].$$

The agent bears all the risk for shortfalls in q, and the principal gets

$$q_i - I_i - B(a^*) - C_{FB}(a^*).$$

Now, suppose the agent bears less risk and takes only a fraction τ of the shortfall in q.

Income in state i becomes

(1) $I_i = \tau q_i - t[B(a^*) - C_{FB}(a^*)]$,

where t is a constant (discussed in detail below) and τ measures the fraction of risk borne by the agent. Proposition 1 gives sufficient conditions for τ being less than one:

PROPOSITION 1: Assume A1-A2 and a risk-neutral agent. If

$$\tau \in \bigcap_{a:a \neq a^* \text{ and } \tau(a) \leq 1} [\tau(a), 1],$$

where

$$\tau(a) = \frac{1}{\beta} \frac{1}{B(a) - B(a^*)} \left\{ (\alpha + \beta \hat{I}) \left(\frac{1}{K(a)} - \frac{1}{K(a^*)} \right) + \frac{G(a^*)}{K(a^*)} - \frac{G(a)}{K(a)} \right\},$$

then there exists an optimal contract paying the agent $I_i = \tau q_i - t [B(a^*) - C_{FB}(a^*)]$ for some value of t.

The proof is straightforward and revealing. To emphasize the underlying logic, I make two simplifying assumptions about utility, both of which are easily generalized. First, I specialize the risk-neutral income utility to $V(I) = I$, rather than to $V(I) = \alpha + \beta I$. Second, I use the additively separable form of utility, setting $U(a,I) = G(a) + V(I)$, or here, $G(a) + I$.

PROOF:

For the optimal action, the principal calculates the least costly method of getting the agent to choose action a^*. The incentive scheme must minimize the principal's expected payment to the agent while still inducing him to act. This is a programming

5

problem, including individual rationality, incentive compatibility, and feasibility constraints:

(P1) $\quad \underset{\{I_i\}}{\text{MIN}} \sum_{i=1}^{n} \pi_i (a^*)$

subject to

(IR) $\quad \sum_{i=1}^{n} \pi_i (a^*)[G(a^*)+I_i] \geq \overline{U},$

(IC) $\quad \sum_{i=1}^{n} \pi_i(a^*) [G(a^*)+I_i] \geq \sum_{i=1}^{n} \pi_i(a^*) [G(a)+I_i]$ for $a \neq a^*,$

(FEAS) $\quad I_i \leq \infty$ for all i.

We now must determine the τ in equation (1) that will satisfy these conditions. This means choosing τ to satisfy

$$\sum_{i=1}^{n} \pi_i(a^*)I_i = \sum_{i=1}^{n} \pi_i(a^*) \{\tau q_i - t[B(a^*) - C_{FB}(a^*)]\} = C_{FB}(a^*),$$

resulting in

(2) $\quad t = \dfrac{\tau B(a^*) - C_{FB}(a^*)}{B(a^*) - C_{FB}(a^*)}.$

By construction, a τ value between zero and one satisfies the individual rationality constraint (IR). Some values of τ also satisfy the incentive compatibility constraint (IC), as I now show. Substituting equation (2) into (1), the incentive scheme becomes

(3) $\quad I_i = \tau q_i - \tau B(a^*) + C_{FB}(a^*).$

6

This makes the incentive compatibility constraint

$$(4) \quad G(a^*) + \sum_{i=1}^{n} \pi_i(a^*)[\tau q_i - \tau B(a^*) - C_{FB}(a^*)] \geq$$
$$G(a) + \sum_{i=1}^{n} \pi_i(a)[\tau q_i - \tau B(a^*) - C_{FB}(a^*)],$$

which simplifies to

$$(5) \quad G(a^*) - G(a) \geq \tau [B(a) - B(a^*)].$$

Whether or not a risk-neutral agent bears all the risk depends on whether there is a gap between $G(a^*)-G(a)$ and $B(a)-B(a^*)$.[2] This gap is not solely a matter of chance, however. The principal chooses a* to maximize $B(a)-C(a)$ or, in the risk-neutral case, $B(a)-C_{FB}(a)$. Since a* is the optimal action, it satisfies $B(a^*)-C_{FB}(a^*) \geq B(a)-C_{FB}(a)$. Rearranging and using the definition of C_{FB}, we have

$$(6) \quad G(a^*) - G(a) \geq B(a) - B(a^*).$$

If the inequality in (6) is strict, τ can be less than one, meaning that the agent does not assume all the risk. There are three cases to consider, depending on the sign of each side of (6).

(i) Both $G(a^*)-G(a)$ and $B(a)-B(a^*)$ are positive. In this case, a has the larger gross payoff but is more costly to implement than a*. Clearly, if (6) holds, any τ in the relevant range of [0,1] satisfies (5).

(ii) If $G(a^*)-G(a)$ is positive and $B(a)-B(a^*)$ is negative, any τ works. In this case, the less costly action, a*, also has the better payoff.

(iii) Both G(a*)-G(a) and B(a)-B(a*) are negative. In this case, a* is more costly but has a better payoff. We usually think of this as the "normal" case. With negative numbers, division reverses signs, so (5) implies that τ, the fraction of risk the agent does bear, can fall anywhere in the interval

$$\tau \in \left[\frac{G(a^*) - G(a)}{B(a) - B(a^*)}, 1\right].$$

With a more general utility function, this becomes the condition stated in the proposition:

$$(7) \quad \tau(a) = \frac{1}{\beta} \frac{1}{B(a) - B(a^*)} \left\{(\alpha + \beta \hat{I})\left(\frac{1}{K(a)} - \frac{1}{K(a^*)}\right) + \frac{G(a^*)}{K(a^*)} - \frac{G(a)}{K(a)}\right\}.$$

Even equation (7) understates the full range of incentive schemes wherein the principal bears risk. With more than two states, the sharing rule need not be linear, and a single-parameter τ will not capture all possible deviations from the classical case. In general, the solution set will be the convex hull of extreme points, a multidimensional "flat" or "face" of the constraint set for the linear programming problem (P1).

III. Convergence

Solutions in which the principal assumes some risk are more than curiosities. As risk aversion approaches zero, the risk borne by the agent converges to a number less than one. The traditional solution offers a poor approximation of this, even near zero.

Unfortunately, I have results only for the two-state case -- the sole case with closed-form solutions for the risk-averse problem. Such strong assumptions seem to be necessary for convergence results. For instance, GH often assume only two states, or negative exponential utility. Without strong restrictions, odd things can happen in the

8

model: The (IR) constraint may not bind, higher profits may mean less money for the agent, or the agent may get more money for less effort.

A. Limiting Cases

With only two states, the single-parameter τ fully describes how much risk the agent bears. Usually, the risk-averse solutions converge to the solution with the smallest τ value (rather than to the classical solution of $\tau=1$). Some exceptions exist because the optimal action can switch at zero, which in turn causes a discrete jump in the risk burden.

To explore convergence, we must first make sure that the utility functions do indeed converge. If we index the income utility function by risk aversion γ, $V(\gamma,I)$, we embody this convergence as a new assumption.

Assumption A3: As ∂ approaches 0, $V(\gamma,I)$ converges uniformly to $\alpha+\beta I$, $\alpha,\beta \neq 0$, on the interval $[-q_n, q_n]$.

Though natural, this assumption does restrict utility functions. For example, the negative exponential function $-e^{-\gamma(I-a)}$ converges to zero, a constant function that is inadmissible by A1.

The statement of proposition 2 requires a little groundwork. First, the proof uses the closed-form solution for the two-state case found by GH:

$$(8) \quad v_1 = \frac{\pi_2(a_j)[\overline{U} - G(a_k)] - \pi_2(a_k)[\overline{U} - G(a_j)] / K(a_j)}{\pi_1(a_k) - \pi_1(a_j)}.$$

$$(9) \quad v_2 = \frac{\pi_1(a_j)[\overline{U} - G(a_k)] - \pi_1(a_k)[\overline{U} - G(a_j)] / K(a_j)}{\pi_2(a_k) - \pi_2(a_j)}.$$

9

The derivation of these formulas depends crucially on GH's proposition 6, which proves that the agent is indifferent between the optimal action a* and some less costly action. Two possibilities can make convergence problematic. As risk aversion falls, either the optimal action or the less costly action may change. A change in the optimal action matters for the convergence result, but whether or not a change in the less costly action does is unclear. I have produced neither a proof nor a counterexample for this case. Thus, the statement of proposition 2 reflects these two possibilities.

Now define the unique profit share for a given utility function and risk aversion as $\tau(V,\gamma)$. Further define the minimum τ in equation (7) as τ_{min}. This gives

PROPOSITION 2: Given assumptions A1-A3, if the optimal action and the indifferent alternative action do not change for risk aversion in the neighborhood of zero, then

$$\lim_{\gamma \to 0} \tau(V, \gamma) = \tau_{min}.$$

As before, to ease the notational burden and emphasize the logic, I present the proof for the additively separable case. The generalization to other utility functions is straightforward.

PROOF:

In the risk-neutral case, we know from equation (6) that $\tau_{min} = \dfrac{G(a_k) - G(a_i)}{B(a_i) - B(a_k)}$, which clearly depends on the optimal action a_k and a particular alternative a_i. This implies an income difference between states of

$$(10) \quad I_2 - I_1 = G(a_j) - G(a_k) \, / \, \pi_1(a_j) - \pi_1(a_k).$$

10

In the limit of the risk averse case, the optimal incomes are given by the limits of equations (8) and (9).

$$(11) \quad I_1 = v_1 = \frac{\pi_2(a_i)\left[\bar{I} - G(a_k)\right] - \pi_2\left[\bar{I} - G(a_j)\right]}{\pi_1(a_k) - \pi_1(a_j)}.$$

$$(12) \quad I_2 = v_2 = \frac{\pi_1(a_i)\left[\bar{I} - G(a_k)\right] - \pi_1(a_k)\left[\bar{I} - G(a_j)\right]}{\pi_2(a_k) - \pi_2(a_j)}.$$

Since $\pi_1(a_j) + \pi_2(a_j) = 1$, we can express the probabilities in terms of $\pi_1(\cdot)$'s. Making this substitution and collecting terms yields

$$I_1 = \frac{\left[\pi_1(a_k) - \pi_1(a_j)\right]\bar{I} + [1 - \pi_1(a_k)]G(a_j) - \left[1 - p_1(a_j)\right]G(a_k)}{\pi_1(a_k) - \pi_1(a_j)}.$$

$$I_2 = \frac{\left[\pi_1(a_j) - \pi_1(a_k)\right]\bar{I} + \pi_1(a_k)G(a_j) - \pi_1(a_k)G(a_k)}{\pi_1(a_j) - \pi_1(a_k)}.$$

Taking the difference and simplifying, we find

$$(13) \quad I_2 - I_1 = G(a_j) - G(a_k) \,/\, \pi_1(a_j) - \pi_1(a_k) = (I_2 - I_1)k,$$

which matches (10).

The equality between equations (10) and (13) proved so far depends on the constancy of both the optimal action and the alternative action. I conjecture that even if the alternative action switches in the neighborhood of zero, the equality (and thus the proposition) still holds.

B. Action Shifts

Proposition 2 does not hold when the optimal action shifts at zero. Suppose one action is best at a risk aversion of zero and another at a risk aversion greater than zero. As the action changes, so too does the sharing rule. The best way to see this is with a simple two-act example. Here, the principal induces the better action at zero risk aversion, but pays a flat fee and accepts the lower action for risk aversion greater than zero.

We start out with $B(a^*)-C(a^*)=B(a)-C(a)$, or indifference between the two actions, so that the switch happens at zero. Notice that this sets τ equal to one, meaning that the agent bears all the risk. We next want $B(a_2)-C(a_2) < B(a_1)-C(a_1)$, making the lower action preferred for $\gamma > 0$. To do this, set $V(I) = I - \gamma I^2$. Then, $h(v) = \left\{ 1 + \sqrt{(1-4\gamma v)} \right\} / 2\gamma$. With $h(v)$ in hand, we can assess the second-best costs once we have calculated v_1 and v_2. The goal is then to show that, in some cases, $\dfrac{\partial C(a_2)}{\partial \gamma} > 0$. If this is true, an increase in γ leads to the principal preferring action a_1, since the cost of action a_2 increases while the rest of the variables, $B(a_2)$, $B(a_1)$, and $C(a_1)$, remain unchanged. ($C[a_1]$ is a fixed payment independent of state.)

Simplifying v_1 and v_2 from equations (8) and (9), we have

$$v_1 = \bar{I} - \gamma \bar{I}^2 + \frac{G(a_1) - G(a_2) + \pi_1(a_1)G(a_2) - \pi_1(a_2)G(a_1)}{\pi_1(a_2) - \pi_1(a_1)}.$$

$$v_2 = \bar{I} - \gamma \bar{I}^2 + \frac{\pi_1(a_2)G(a_1) - \pi_1(a_1)G(a_2)}{\pi_1(a_2) - \pi_1(a_1)}.$$

The last terms in each of these expressions are constant with respect to γ, so we may rewrite them as

$$v_1 = \bar{I} - \gamma \bar{I}^2 + P.$$

12

$$v_1 = \quad \bar{I} - \gamma\bar{I}^2 + P.$$

$$v_2 = \quad \bar{I} - \gamma\bar{I}^2 + Q.$$

Using the above equations, we can solve for I_1 and I_2 and thus for $C(a_2)$:

$$I_1 = \quad \frac{1}{2\gamma} - \frac{1}{2\gamma}\left[1 - 4\gamma\left(\bar{I} - g\bar{I}^2 + P\right)\right]^{\frac{1}{2}}.$$

$$I_2 = \quad \frac{1}{2\gamma} - \frac{1}{2\gamma}\left[1 - 4\gamma\left(\bar{I} - g\bar{I}^2 + Q\right)\right]^{\frac{1}{2}}.$$

Notice that $\partial I_1/\partial\gamma$ and $\partial I_2/\partial\gamma$ have the same sign, matching $\partial C(a_2)/\partial\gamma$. Explicitly calculating the first of these derivatives, we have

$$\partial I_1 / \partial\gamma = \frac{1}{\gamma}[1 - 4\gamma(v_1)]^{\frac{1}{2}}\left[v_1 - \bar{I}^2\right].$$

The first two terms are positive, while the last can be rewritten as $\left[\bar{I} - (1+\gamma)\bar{I}^2 + P\right]$. As $\gamma \to 0$, the last term approaches $\bar{I} - \bar{I}^2 + P$. For $\bar{I}$ not too large, that term is positive, and we have the counterexample.

In this counterexample, the agent bears all the risk if he is risk neutral, but assumes none at all if he is even slightly risk averse. In other words, convergence fails in a spectacular way. But it may fail in more prosaic fashions as well. The limit of the risk-averse case may be higher or lower than τ_{min}. Figure 1 schematically illustrates these possibilities.

Mathematically, convergence fails because of a difference between τ_{min} for the risk-neutral case and τ for the limit of the risk-averse case. This difference is

$$(14) \quad \tau_{min} - \lim_{\gamma\to 0}\tau = [G(a_1) - G(a_j)] / [B(a_j) - B(a_1)] - [G(a_k) - G(a_j)] / [B(a_j) - B(a_k)].$$

13

Indifference at zero risk aversion implies $G(a_1) + B(a_1) = G(a_k) + B(a_k)$, leaving open two distinct possibilities: Either a_k or a_1 can be the high-cost, high-benefit action. If $B(a_k) > B(a_1)$, then $G(a_k) < G(a_1)$, and vice versa. The sign of equation (14) then can go either way.

C. Increasing the Number of Actions

The lowest share of risk the agent can take, τ_{min}, is decreasing in the gap in the principal's payoff between the chosen and the indifferent act, $[G(a_k) - G(a_j)]/[B(a_j) - B(a_k)]$. It seems intuitive that as the number of actions increases, the gap decreases and hence τ_{min} moves toward one, its value in the continuous action case. But it is possible to work the convergence so that exceptions occur. If B and G are continuous functions, some condition on the difference, such as $\lim |a_k - a_k + 1| = 0$, would ensure the result.

IV. Conclusion

The traditional solution to the risk-neutral principal-agent problem is misleading. With finite states and finite actions, many solutions exist, and in all but one of these the *principal* bears the risk. The traditional solution cannot even claim to be the limiting case as risk aversion decreases: In fact, it is the solution farthest away from the limit.

These results have two main consequences. First, they caution us against using the traditional solution as an approximation for the less tractable risk-averse case. This explains the divergence between Haubrich's (1991) findings and those of Jensen and Murphy (1990). Second, they also illustrate the range, power, and tractability of GH's version of the principal-agent model.

Nevertheless, the results presented here should be taken as preliminary -- brief observations of a rare nocturnal animal. Theorem 1 provides sufficient, but not necessary, conditions for multiple solutions and does not characterize all possible solutions. The convergence results require even stronger restrictions and depend on the two-act case.

14

15

Still, I believe the scattered sightings reported here show a surprising -- and noteworthy --
aspect of the principal-agent model.

FOOTNOTES

1. Although GH do not consider multiple solutions in the risk-neutral case, they are quite careful in stating their theorems. Hence, this result does not imply any error in their work.

2. Haubrich (1991) provides several numerical examples of problems of this type, showing that solutions do exist and that the theorem is not vacuous.

REFERENCES

Grossman, Sanford J., and Oliver D. Hart, "An Analysis of the Principal-Agent Problem," *Econometrica*, vol. 51, no. 1, January 1983, pp. 7-45.
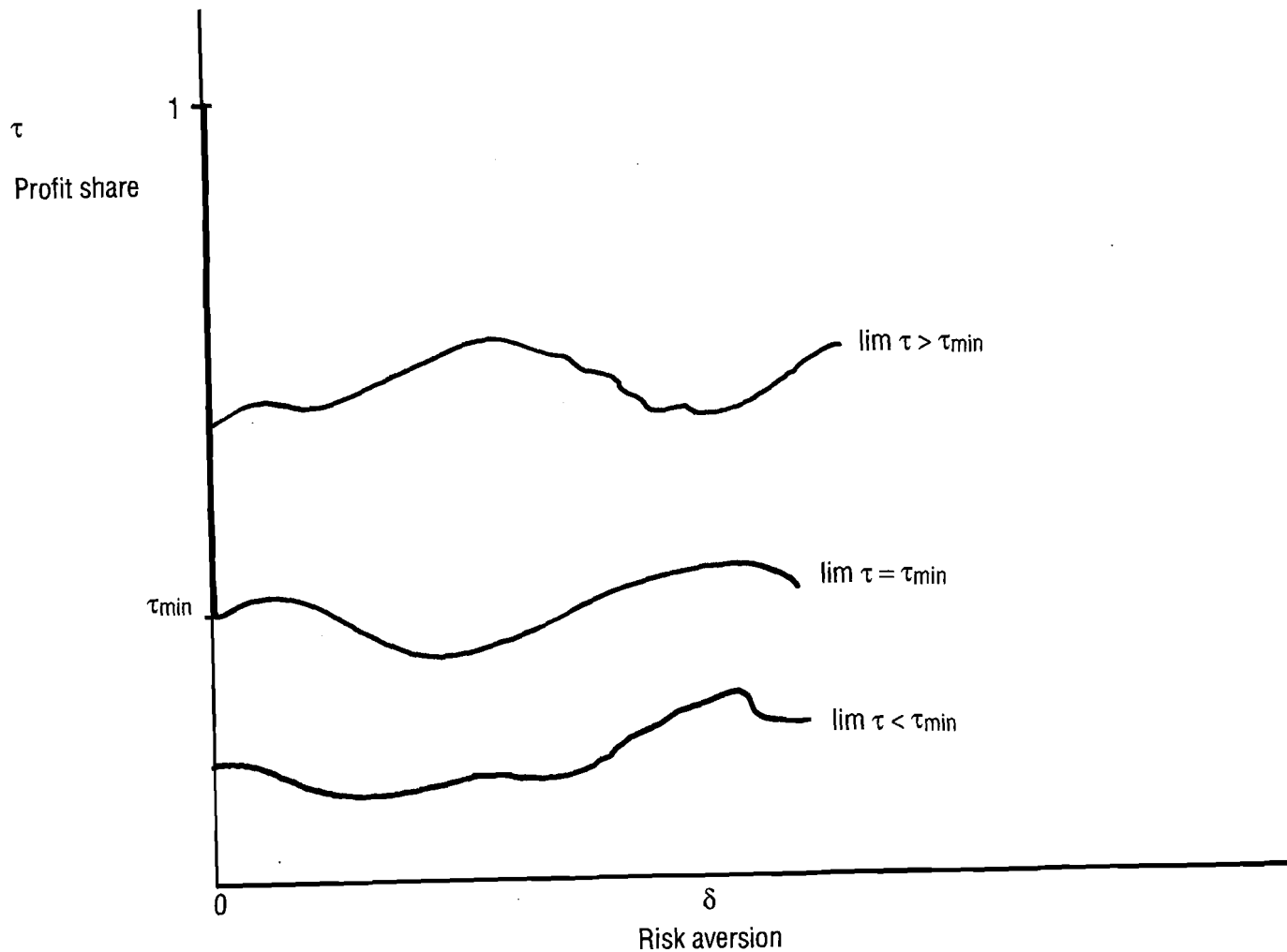
Haubrich, Joseph G., "Risk Aversion, Performance Pay, and the Principal-Agent Problem," Federal Reserve Bank of Cleveland, Working Paper No. 9118, December 1991.

Holmstrom, Bengt, "Moral Hazard and Observability," *Bell Journal of Economics*, vol. 10, 1979, pp. 74-91.

Jensen, Michael C., and Kevin J. Murphy, "Performance Pay and Top-Management Incentives," *Journal of Political Economy*, vol. 98, no. 2, April 1990, pp. 225-264.

Mirrlees, James A., "The Optimal Structure of Incentives and Authority within an Organization," *Bell Journal of Economics*, vol. 7, 1976, pp. 105-131.

Figure 1
Convergence of Sharing Rule

Source: Author.