Vol. 37, No. 1

ECONOMIC REVIEW

2001 Quarter 1

FEDERAL RESERVE BANK
OF CLEVELAND

http://clevelandfed.org/research/review/

ECONOMIC REVIEW

2001 Quarter 1
Vol. 37, No. 1

In the standard solution to the principal–agent problem, a risk-neutral agent bears all the risk. The author shows that, in fact, multiple solutions exist, and often the risk-neutral agent is not the sole bearer of risk. 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 result in agents bearing significantly less risk than the standard solution suggests.

What rule should a central bank interested in inflation stability follow? Because monetary policy tends to work with lags, it is tempting to use inflation forecasts to generate policy advice.This article, however, suggests that the use of forecasts to drive policy is potentially destabilizing. The problem with forecast-based policy is that the economy becomes vulnerable to what economists term "sunspot" fluctuations.These welfare-reducing fluctuations can be avoided by using a policy that puts greater weight on past, realized inflation rates rather than forecasted, future rates.

Economic Review is published quarterly by the Research Department of the Federal Reserve Bank of Cleveland. To receive copies or to be placed on the mailing list, e-mail your request to *4d.subscriptions@clev.frb.org* or fax it to 216/579-3050.

Economic Review is available electronically through the Cleveland Fed's site on the World Wide Web: *http://www.clev.frb.org.*

We welcome your comments, questions, and suggestions. E-mail us at *editor@clev.frb.org.*

ISSN 0013-0281

Sharing with a Risk-Neutral Agent

by Joseph G. Haubrich

Joseph G. Haubrich is an economist and consultant at the Federal Reserve Bank of Cleveland.
He thanks Patricia Beeson and Yukiko Hirao for helpful comments.

Introduction

In the classical principal–agent problem, a risk-neutral agent bears all the risk. This particular solution, while acknowledged as a special case, is prominent in the minds of economists because the more general risk-averse case does not easily yield numerical results. For example, Jensen and Murphy (1990) find a divergence between the risk actually borne by chief executive officers and the risk-neutral solution, which seems too large to be accounted for by reasonable levels of risk aversion.

Although the standard risk-neutral solution is correct, it is also misleading. Other solutions exist wherein the agent does not bear all the risk, and these may be considered more "natural," since they are limits of the risk-averse case.

Specifically, in Grossman and Hart's (1983) principal–agent problem, where there is 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, 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. Because less risk makes a risk-averse agent happier, he demands a lower risk premium, in turn making the principal happier. But exceptions occur. There are 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) differs substantially when the principal wants to induce distinctly different actions.

By increasing the number of actions, the 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 misleading to compare actual contracts in which risk aversion is important—executive compensation, for instance—with the predictions of the risk-neutral principal–agent model. Stated more positively, these results show how principal–agent theory implies the relatively flat sharing rules that are observed in practice.

■ **1** Although Grossman and Hart do not explicitly mention multiple solutions in the risk-neutral case, they are careful in stating their theorems. Hence, this result does not imply any error in their work.

I. Sharing Rules

The Model

First let us quickly review the assumptions, notation, and approach of the Grossman–Hart model. For concreteness, assume the principal owns a firm, but she delegates its management to the agent. There is a finite number of outcomes (gross profit states), $q_1 < q_2 < \ldots < q_n$. The principal, who is risk neutral, cares only 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*, which the principal cannot observe. The principal does observe the outcome, however, and, like the agent, knows that different actions determine the probability of the outcome states. Both know $\pi_i(a)$, the probability of outcome q_i 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 $A=\{a_1, a_2, a_3 \ldots a_m\}$, making the principal's expected benefit from an action equal to

$$B(a) = \sum_{i=1}^{n} \pi_i(a) q_i .$$

To avoid the problem of increasingly larger penalties being imposed with progressively smaller probabilities (see Mirrlees [1976]), assume that $\pi_i(a)$ is strictly greater than zero for all states and actions.

The agent likes income, but he dislikes effort. His utility function, $U(a,I)$, depends positively on his income from the principal, I, and negatively on his action, a. Grossman and Hart 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, concave function with the domain $[\underline{I},\infty]$ and $\lim_{I \to \underline{I}} = -\infty$. G and K are real-valued, continuous functions defined by A, and K is strictly positive. 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 $\bar{U}$, that is, the expected utility he can achieve working elsewhere. Sometimes this is derived from an outside income $\bar{I}$, so that $\bar{U} = V(\bar{I})$. If the principal does not offer him a contract worth at least $\bar{U}$, the agent will take another job. To make the model at all interesting, some income level should induce the agent to work. Grossman and Hart formalize this as

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

To see what happens when this assumption does not hold, consider the negative exponential utility $-e^{-k(I-a)}$ and $\bar{U} = 5$. In this case, even infinite income could not make the agent work.

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

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

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

As Grossman and Hart put it, "$C_{FB}(a)$ 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, $I_i = \{I_1, I_2, \ldots I_n\}$, wherein payment I_i depends on the observed final state, q_i. Given this, the agent will choose the 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 will induce 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 $B(a)-C(a)$.

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. With a risk-averse agent, the principal minimizes the agent's risk, subject to meeting the incentive constraints. For a risk-neutral agent, only the incentives matter, and any incentive-compatible risk configuration will work. 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.

The traditional solution assigns all risk to the agent, who delivers a fixed payment to the principal.

The agent, then, receives

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

The agent bears all the risk for shortfalls in q, and the principal gets her expected benefit.

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

Now, suppose the agent bears less risk and takes only a fraction of the shortfall in q. Income in state i becomes

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

where t is a constant 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

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

then an optimal contract exists that pays the agent $I_i = \tau q_i - t\,[B(a^*) - C_{FB}(a^*)]$ for some value of t. This somewhat complicated condition guarantees there is a "gap" or "jump" between the principal's payoff in different states.

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

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 problem, including individual rationality (IR), incentive compatibility (IC), and feasibility constraints (FEAS).

$$\text{(P1)}\qquad \text{MIN} \sum_{i=1}^{n} \pi_i(a^*),$$

subject to

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

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

$$\text{(FEAS)}\qquad I_i \leq \infty \text{ for all } i.$$

We now must determine the value of 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

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

By construction, values between 0 and 1 satisfy the individual-rationality constraint. Some values of τ also satisfy the incentive-compatibility constraint, as I will now show. Substituting equation (2) into equation (1), the incentive scheme becomes

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

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 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] But this gap is not solely a matter of chance: The principal chooses a^* to maximize $B(a) - C(a)$ or, in the risk-neutral case, $B(a) - C_{FB}(a)$. Because 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

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

2 Haubrich (1994) provides several numerical examples of problems of this type, showing that solutions do exist and the theorem is not vacuous.

If the inequality in equation (6) is strict, τ can be less than one, meaning the agent does not assume all the risk. There are three cases to consider, depending on the sign of each side of equation (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 equation (6) holds, any τ in the relevant range of [0,1] will satisfy equation (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 equation (5) implies that τ, the fraction of risk borne by the agent, 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:

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

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 classic 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).

II. 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.

The convergence results are for the two-state case—the sole case with closed-form solutions for the risk-averse problem. Answering convergence questions usually requires strong assumptions. For instance, Grossman and Hart assume only two states, or negative exponential utility. Without strong restrictions, odd things can occur in the model: The individual-rationality constraint may not bind, higher profits may mean less money for the agent, or the agent may get more money for less effort.

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 value (rather than the classic solution of τ=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 the utility functions do, in fact, converge. If we index the income utility function by risk aversion γ, (γ, I), we embody this convergence as a new assumption.

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

Although it is 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 assumption (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 Grossman and Hart:

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

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

The derivation of these formulas depends crucially on Grossman and Hart's proposition 6, which proves the agent is indifferent between the optimal action a^* and some less costly action. The existence of two possibilities makes 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 it is not clear whether a change in the less costly action makes a difference. I have produced neither a proof nor a counterexample for this case. Thus, the statement of proposition 2 reflects these two possibilities.

The unique profit share for a given utility function and risk aversion is defined as $\tau(V,\gamma)$, and the the minimum in equation (7) is defined as $\tau_{\min}$.

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

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

To ease the notational burden and to 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

$$\tau_{\min} = \frac{G(a_k) - G(a_i)}{B(a_i) - B(a_k)},$$

which clearly depends on the optimal action a_k and a particular alternative a_i. This implies an income difference between states of

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

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

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

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

Because $\pi_1(a_j) + \pi_2(a_j) = 1$, the probabilities can be expressed in terms of $\pi_1(\bullet)$. Making this substitution and collecting terms yields

$$I_1 = \{[\pi_1(a_k) - \pi_1(a_j)]\,\bar{I} + [1 - \pi^1(a_k)]\,G(a_j) - [1 - p_1(a_j)]G(a_k)\}\,\{\pi_1(a_k) - \pi_1(a_j)\}^{-1}$$

$$I_2 = \{[\pi_1(a_j) - \pi_1(a_k)]\,\bar{I} + \pi_1(a_k)G(a_j) - \pi_1(a_k)G(a_k)\}\,\{\pi_1(a_j) - \pi_1(a_k)\}^{-1}$$

Taking the difference and simplifying, we find

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

which matches equation (10).

The equality between equations (10) and (13) 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.

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 does the sharing rule. The best way to illustrate this is 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 begin with $B(a^*)–C(a^*)=B(a)–C(a)$, or indifference between the two actions, so that the switch occurs at zero. This sets τ equal to one, meaning 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, $h(v) = \sqrt{[1 + (1 - 4\gamma v)]}\,/2\gamma$. With $h(v)$ in hand, we can assess the second-best costs once we have calculated v_1 and v_2. The goal is to show that, in some cases, $\partial C(a_2)/\partial\gamma > 0$. If this is true, an increase in leads the principal to prefer 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. ($C[a_1]$ fixed payment independent of state.)

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

$$v_1 = \bar{I} - \gamma\bar{I}^2 + [G(a_1) - G(a_2) + \pi_1(a_1)G(a_2) - \pi_1(a_2)G(a_1)]\,[\pi_1(a_2) - \pi_1(a_1)]^{-1}$$

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

This represents a shift in the optimal action induced by the principal. Both actions remain feasible. The last terms in each of these expressions are constant with respect to γ, so we may rewrite them as

$$v_1 = \bar{I} - \gamma\bar{I}^2 + P$$
$$v_2 = \bar{I} - \gamma\bar{I}^2 + Q.$$

and solve for I_1, I_2, and $C(a_2)$:

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

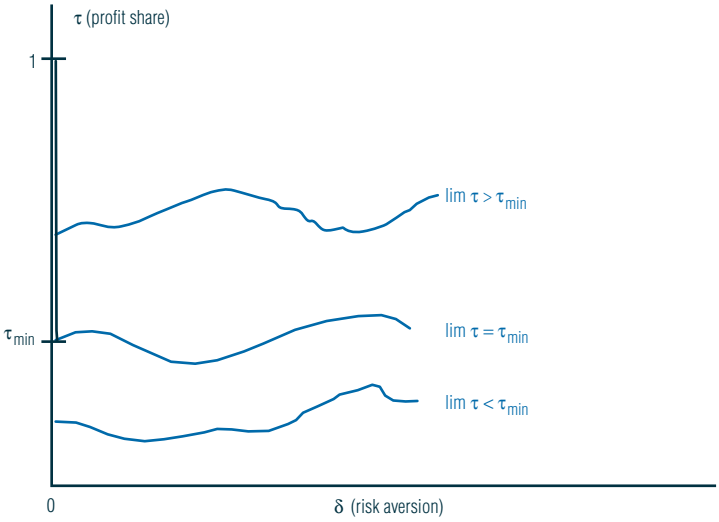
$$I_2 = \frac{1}{2\gamma} - \frac{1}{2\gamma}\left[1 - 4\gamma(\bar{I} - g\bar{I}^2 + P)\right]^{\frac{1}{2}}.$$

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

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

FIGURE 1

Convergence of the Sharing Rule



SOURCE: Author's calculations.

The first two terms are positive, while the last can be rewritten as $[\bar{I} - (1+\gamma)\bar{I}^2 + P]$. As $\gamma \to 0$, the last term approaches $\bar{I} - \bar{I}^2 + P$. For values of $\bar{I}$ that are 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 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 $\tau_{\min}$. Figure 1 schematically illustrates these possibilities.

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

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

Indifference at zero risk aversion implies $G(a_1) + B(a_1) = G(a_k) + B(a_k)$, creating 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.

Despite the myriad possibilities, nonconvergence remains a special case. To start with, the principal must be indifferent between two different actions of the risk-neutral agent, and she must strictly prefer one action for arbitrarily small levels of risk aversion.

Increasing the Number of Actions

The lowest share of risk the agent can take, $\tau_{\min}$, 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 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.

III. 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 of the less tractable risk-averse case. This may explain why Jensen and Murphy (1990) found CEOs bearing a surprisingly low amount of risk. It also explains, in part, why the numerical calculations of CEO risk in Haubrich (1994) were so small, even for very low levels of risk aversion. Second, they illustrate the range, power, and tractability of Grossman and Hart's version of the principal–agent model.

Nevertheless, the results presented here should be taken as preliminary—brief observations of a rare nocturnal animal. Proposition 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. Still, I believe the scattered sightings reported here show a surprising—and noteworthy—aspect of the principal–agent model.

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," *Journal of Political Economy,* vol. 102, no. 2 (April 1994), pp. 258–76.

Holmstrom, Bengt. "Moral Hazard and Observability," *Bell Journal of Economics*, vol. 10, no. 1 (Spring 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. 225–64.

Mirrlees, James A. "The Optimal Structure of Incentives and Authority within an Organization," *Bell Journal of Economics,* vol. 7, no. 1 (Spring 1976), pp. 105–31.

Monetary Policy and Self-Fulfilling Expectations: The Danger of Forecasts

by Charles T. Carlstrom and Timothy S. Fuerst

Charles T. Carlstrom is a senior economic advisor at the Federal Reserve Bank of Cleveland; Timothy S. Fuerst is a professor of economics at Bowling Green State University.

Introduction

Economists have long argued that the best, surest way for a central bank to do its job is to adopt some sort of rule and stick to it. They reason that a central bank's short-term objectives may be inconsistent with its long-term goals; a rule should prevent the bank from undermining long-range goals for the sake of short-term results.[1]

Using a rule also makes a central bank's actions more transparent. Consider the case of a central bank whose long-term goal is inflation stability. If the bank doesn't specify its long- and short-term objectives, the public is apt to misinterpret the bank's actions, making inflation stability difficult to achieve. For example, to enhance the market's functioning over the business cycle, a central bank may make a temporary change, which the public may confuse with a change in the bank's long-term objective. If no explicit reasons for the bank's actions are given, public expectations about future inflation have no moorings.

But what kind of rule is best for a central bank that wants to achieve stable inflation? One of the earliest, most famous proposals was Milton Friedman's constant-money-growth rule. He argued that the monetary authority should ignore short-run considerations altogether, because attempts to stabilize inflation—or even output—would ultimately make matters worse. Long and variable amounts of time pass before changes in the money supply affect prices, so monetary authorities cannot be sure when and to what degree their policies take effect. This, ironically, means that stabilization policies would potentially be destabilizing. Milton Friedman concluded that the monetary authority should just commit to expanding the money supply by a constant amount every year.

The chronic, widespread instability in money demand that has been apparent to many economists since at least the mid-1970s weakened this position, and the unexpected shift in money demand in the early 1990s signaled its demise. The relationship between money and prices seems less predictable now than it once did. Most policymakers now recognize that a constant-growth rule will not prevent inflation from varying substantially over short and long periods. A growing number of central banks recently have moved toward the idea that they should target the inflation rate directly. For example, Canada, the United Kingdom, Sweden, and New Zealand have all adopted explicit inflation targets.

Inflation targeting, however, is an objective, and the best way to achieve it remains controversial. That is, would it be better to respond proactively to stop inflation before it increases or to respond only after

1 This refers to the advantages of using rules because of the time-inconsistency problem. See, for example, Kydland and Prescott (1977) and Barro and Gordon (1983).

realized inflation has crept up? While there is no universal agreement on the best policy rule to stabilize inflation, central banks with inflation targets have found it necessary to base policy changes on inflation projections. In New Zealand's case, these projections are set two to three years ahead. In the United Kingdom, they have a current target of 2.5 percent and forecast inflation two years ahead in setting policy. In fact, the central bank of New Zealand states that its "inflation projections relative to the inflation target range are the critical input in the quarter by quarter formulation of monetary policy."[2]

If monetary policy is to stabilize the inflation rate over any but the longest time horizon, then the monetary authority must look ahead and respond to what inflation is expected to be. This is crucial, given the long lags between monetary policy and price changes. Without such forward-looking, pre-emptive behavior, the monetary authority is repeatedly responding to past inflation shocks, many of which are temporary, with no bearing on future inflation. The result may be unnecessarily wide price swings.[3]

Although the United States has no official price level target, it is clearly committed to not letting inflation accelerate. Consequently, we rely on forecasts. As Alan Greenspan commented: "Implicit in any monetary policy action or inaction is an expectation of how the future will unfold, that is, a forecast."[4]

Inflation targets are meant to reduce uncertainty and tie down expectations. This paper argues, however, that far from pinning down expectations, such a policy leaves expectations completely free.[5] The danger in basing monetary policy on forecasts is that it creates a situation in which policy depends on expected inflation and expected inflation, in turn, depends on policy.[6] As a result, nothing pins down either one. The "anchor" that inflation targeting is supposed to provide may well be illusory, leaving monetary policy (and consequently real output) without any moorings.

The consistent use of inflation forecasts, which is necessary with strict inflation targeting, leaves the system vulnerable to self-fulfilling inflation expectations. The pernicious event that triggers these self-fulfilling cycles is known as a *sunspot*. We argue that instead of using inflation forecasts in conducting monetary policy, thereby creating the potential for sunspot-induced volatility, the monetary authority should respond aggressively to *past* inflation. Although inflation can never be truly stable if only past inflation is responded to, only in this way will monetary policy truly provide the anchor that pins down inflation expectations.

I. Sunspots and Lack of Coordination

The possibility of sunspot-induced, self-fulfilling expectations can arise if monetary policy depends on what the public is expected to do, and the public, in turn, bases its behavior on policy actions. This can lead to the well-known problem of "infinite regress," in which the public's behavior and monetary policy affect each other in turn, and there is nothing objective on which to "pin down" either. Outcomes are determined by each side's beliefs about what the other is expected to do.

A simple noneconomic example illustrates this possibility. Suppose Chuck's decision about whether to attend a party depends on whether he expects Tim to go; Tim's decision, in turn, depends on whether he expects Chuck to be there. Now suppose that Tim believes Chuck will go to the party only if it rains in Tahiti. Tim's belief will be self-fulfilling: If it rains in Tahiti, Tim will go to the party (because he expects Chuck to), and so will Chuck (because Tim is going).

Economists refer to this as "sunspot" behavior. An event is called a sunspot if it affects some economic variable (such as inflation) only because the public believes it does. A sunspot is therefore purely extraneous information (for example, rain in Tahiti) that affects behavior because it leads to a circle of self-fulfilling expectations. If the public expects prices to be higher tomorrow, it acts on this belief, setting in motion a series of forces that actually cause prices to rise.

Perhaps the most famous economic example of self-fulfilling expectations is that of the bank runs during the Great Depression. Because of the first-come-first-served rule, it was in depositors' best interest to withdraw their money whenever they thought the bank might be in financial jeopardy. But if everyone thought the bank was in financial trouble, the ensuing run on the bank would itself cause this trouble. The reason is that much of a bank's portfolio is tied up in assets that cannot be easily liquidated, so that a bank run—or even the

2 See Huxford and Reddell (1996).

3 To quote the Bank of Canada: "There are lags of a year to 18 months or more between monetary policy changes and their effects on inflation and the economy. A chain of events is set in motion that affects consumer spending, sales, production, employment, and other economic indicators. This means that monetary policy must always be forward-looking." See <http:\\www.bankofcanada.ca/en/backgrounds/bgp1.htm>.

4 Greenspan (1994).

5 Because this is a rational-expectations model, we use "forecast" and "expected inflation" interchangeably.

6 Sherwin (1997) writes that because of New Zealand's inflation targeting, "the [central bank] is more likely to be in a position of validating market moves, rather than driving them overtly."

rumor of one—would be a self-fulfilling prophecy. Deposit insurance was instituted to eliminate this.

Self-fulfilling expectations usually occur when multiple players' actions depend on each other but cannot be coordinated. There would be no problem if Chuck and Tim could coordinate their decision about whether to attend the party. Similarly, lack of coordination was crucial in the bank-run problem. The possibility of a disastrous run on an otherwise healthy bank would have been eliminated had depositors been able to coordinate their actions before deciding whether to clean out their accounts. Knowing that others were contemplating withdrawals only out of fear that everyone else would do so would have removed depositors' need to withdraw their funds.

II. Self-Fulfilling Expectations and Monetary Policy

In monetary policy, the two agents that lead to self-fulfilling prophecies are the monetary authority and the public. Such prophecies have become more likely because central banks around the world have found it in their best interest to operate off interest rates (in this country, the federal funds rate). Because the interest rate is the chosen policy instrument, the money supply (inflation's primary determinant) is no longer controlled directly by the monetary authority. It is supplied at whatever level is necessary to achieve the interest rate target. Hence, the potential problem with this approach is that changes in public expectations will indirectly influence money growth, which directly affect (and may even justify) these expectations.

Although our focus is on the response to expected inflation, the basic problem arises under the more extreme assumption of a pure funds rate peg, where monetary policy promises to keep the nominal interest rate constant. Suppose prices today increase. This lowers real money balances, putting upward pressure on nominal interest rates. To keep interest rates constant, the monetary authority must increase the money supply to accommodate the price increases. But at the end of this cycle, real money balances (and hence interest rates) are back where they started.[7] In this example, the central bank's promise to keep interest rates constant obliged it to increase money whenever prices rose.

Thus, prices and nominal money are not pinned down. This indeterminacy is related to the classical statement of monetary neutrality, in which one-time money-supply changes have no real effect, because all dollar prices would respond by the same proportion. With an interest rate peg, the nominal money supply is free in each period because money is supplied at whatever level is needed to keep the peg. Sunspot events can lead to these unanticipated changes in the money supply and prices, but they have no real effect because there is monetary neutrality. Using sunspots is like letting a roulette wheel determine how many zeros should be added (or subtracted) every period from money and prices.

This classical indeterminacy is referred to as a purely nominal indeterminacy. It affects prices and all nominal quantities but affects neither expected inflation nor any real variable. But nominal indeterminacy, coupled with a nominal rigidity, leads to a situation in which expected inflation (and thus real economic variables) is not pinned down.

Suppose that prices are fixed today having been set one period in advance. Now consider a sunspot-induced increase in expected inflation, which puts upward pressure on nominal interest rates (as nominal rates include an expected inflation premium). This obliges the monetary authority to increase the money supply in order to keep nominal interest rates constant. Since prices are fixed today, firms respond to increased money by raising their prices tomorrow. This completes the circle. An increase in expected inflation causes prices to increase tomorrow. Furthermore, because prices were fixed today, the increase in money today would have stimulated real output. Thus not only is expected inflation not anchored, but real output is also without moorings.[8]

A pure funds rate peg, therefore, would make money supply and prices vulnerable to random sunspot events. In principle, nothing governs the size of these sunspots, so prices could be quite volatile and the costs of sunspots quite large.

Of course, the Federal Reserve does not maintain a pure funds rate peg. Instead, changes in inflation (whether past or future) enter heavily into its policy decisions. The question asked in this paper is whether it is better to be proactive and use a forward-looking rule, raising the funds rate when inflation is expected to increase, or to use a backward-looking rule, responding after prices start to rise. For simplicity, we consider rules in which policy responds only to inflation and not to output.[9] We argue that only with an aggressive, primarily backward-looking rule is indeterminacy not a problem. This timing difference mitigates the coordination difficulty because the monetary authority does not "move" until long after the public does.

7 This result is due to Sargent and Wallace (1975). For a general equilibrium analysis, see Carlstrom and Fuerst (1995), which shows that this nominal indeterminacy becomes real in a limited-participation model.

8 The formal details of this logic will be spelled out below.

9 Responding to future output will be similar to responding to future inflation, making self-fulfilling expectations more likely to occur. Responding to past output, however, will make them less likely.

To show this, we develop a simple economic model,[10] first considering a purely flexible-price economy and analyzing the conditions in which nominal indeterminacy will arise, then demonstrating that this nominal indeterminacy will become real when prices are sticky. We conduct the analysis in a perfect-foresight environment because, as is well known, a necessary and sufficient condition for indeterminacy and sunspot fluctuations in a rational-expectations environment with shocks is for there to be indeterminacy in the corresponding-perfect foresight model (without shocks).[11]

III. A Flexible-Price Model

Consider a model economy consisting of numerous infinitely lived households with preferences over consumption, c_t, and disutility over hours worked, L_t. For simplicity, we restrict per-period utility to $U(c, T–L) = \ln(c_t) – L_t$, where T = total time endowment. To reflect people's preference for today rather than tomorrow, utility is assumed to be discounted over time at a constant rate $\beta \equiv 1/(1+\rho) < 1$. Households maximize the infinite discounted value of per-period utilities.

Firms produce the consumption good by using labor supplied by the household according to the simple linear production function, $c_t = f(L_t) = L_t$. To buy the consumption good, households must first acquire money, M_t. Thus, we assume the cash-in-advance (CIA) constraint, $M_t = P_t c_t$.[12] The importance of this assumption is that cash that is being held cannot be invested where it would earn a nominal rate of return, i_t. The nominal interest rate is therefore the opportunity cost of holding money, while the benefit of holding money comes from the consumption it provides.

Monetary policy is assumed to operate off an interest rate objective, where interest rates increase whenever some average of past and future inflation increases.

$$(1)\ \hat{i}_t = \tau[\alpha \hat{\pi}_{t-1} + (1-\alpha)\hat{\pi}_{t+1}],$$

where $\hat{i}_t = i_t - i_{ss}$, $\hat{\pi}_t = \pi_t - \pi_{ss}$, $i_{ss} = \rho + \pi_{ss}$,

where i_t ($\pi_t \equiv \frac{P_t}{P_{t-1}} - 1$) denotes the nominal interest rate (inflation rate) at time t, i_{ss} (π_{ss}) is the steady-state or long-run nominal interest rate (inflation rate). The "hats" thus denote deviations from steady state, and ρ is the fixed, steady-state, real interest rate. According to the policy rule in equation (1), the nominal interest rate increases (decreases) from its long-term trend whenever inflation is higher (lower) than its long-term trend. Monetary policy is said to be aggressive (passive) if $\tau > (<) 1$. The money growth process that supports these rules is endogenous and can be backed out of the CIA constraint. Two extreme forms of this rule are $\alpha = 1$, in which the monetary authority follows a pure backward-looking interest rate rule, and $\alpha = 0$, in which the monetary authority follows a pure forward-looking interest rate rule.

An equilibrium for this economy (see appendix 1) consists of equation (1) and

$$(2a)\ \frac{U_c(t)}{P_t} = \beta R_t \frac{U_c(t+1)}{P_{t+1}}, \text{ where } U_x = \frac{\partial U}{\partial x}$$

$$(3a)\ \frac{U_L(t)}{U_c(t)} = \frac{f_L(t)}{R_t}$$

$$(4)\ M_t = P_t f(L_t).$$

Equation (2a) is the standard Fisher equation. The left side shows the utility lost by forgoing \$1 and hence $1/P$ fewer units of consumption today. The right side shows how much (in terms of utility) is gained tomorrow by investing that dollar bill and earning $R = 1+i$ dollars, which buys R/P units of consumption and provides $U_c(t+1)$ worth of enjoyment tomorrow. On the margin, these two must be equal.

Equation (3a) is the marginal condition for labor. The left side shows how much utility (measured in terms of the consumption good) one loses from working one more hour, while the right side shows the marginal productivity of labor, or how much more consumption one gets by working one more hour.

Because this is a monetary economy, this labor market expression differs from the textbook condition. Defining $(1–t_w) = 1/R$, we see that the nominal (gross) interest rate distorts the economy just as a wage (or, equivalently, a consumption tax) of t_w would. The nominal interest rate acts like a consumption tax because households must acquire cash before buying the consumption good (constraint 4); this has an opportunity cost of R in terms of forgone interest. This, in turn, is equivalent to a wage tax, because labor income is used to purchase consumption. This monetary distortion has an important effect on the existence of sunspot equilibria when one assumes that the central bank conducts policy according to a nominal interest rate rule like the celebrated Taylor rule (Taylor 1993).

10 The model contained in this paper is extremely simple, although the results are quite general. Carlstrom and Fuerst (1999) present technical details in more general environments.

11 Farmer (1999) offers a useful discussion of these issues.

12 A CIA constraint implies that the interest elasticity of money demand is zero and that the velocity of money is unity. Similar results arise in a more general money-in-the-utility-function framework containing today's money balances. See Carlstrom and Fuerst (1998, 1999).

It will be simpler to work in log deviations, so we rewrite (2a) and (3a), plugging in the assumed functional forms, as

(2b) $\hat{i}_t = \hat{\pi}_{t+1} + (\hat{c}_{t+1} - \hat{c}_t)$ and

(3b) $\hat{i}_t = -\hat{c}_t$.

Equation (2b) states that the nominal interest rate consists of an inflation component, $\hat{\pi}_{t+1}$, and a real component. The real interest rate, $\hat{i}_t - \hat{\pi}_{t+1}$, increases whenever consumption is expected to rise over time. Because households prefer a constant consumption stream to a variable one, whenever tomorrow's consumption is expected to be greater (less) than today's, households will tend to borrow (save) more to smooth out their consumption, thus exerting upward (downward) pressure on interest rates.

Eliminating (3b) by substituting out consumption, we obtain[13]

(5a) $\hat{i}_{t+1} = \hat{\pi}_{t+1}$.

This expression, along with the monetary policy rule (1), will determine whether self-fulfilling prophecies are possible. We first analyze the case in which monetary policy looks ahead and show that doing so makes it vulnerable to sunspot-induced fluctuations.

Forward-Looking Interest Rates ($\alpha = 0$)

Suppose the central bank conducts policy according to the forward-looking rule, $\alpha = 0$ in equation (1). Substituting this into (5a) yields

(6) $\hat{\pi}_{t+2} = \left(\frac{1}{\tau}\right)\hat{\pi}_{t+1}$.

Two observations are in order. First, expression (6) starts with π_{t+1} (expected inflation between today and next period), so that the current price level is always free for all values of τ, that is, $\pi_t \equiv \frac{P_t}{P_{t-1}} - 1$ is completely free (P_{t-1} is predetermined by history). This condition does not affect real behavior at this stage and is exactly the analogue of the nominal indeterminacy with pegged interest rates discussed above. Second, the path of expected inflation and thus real behavior is determinate if and only if the mapping in (6) is explosive.[14] Hence, there is real determinacy if and only if $\tau < 1$.

Where does this indeterminacy come from? By definition, indeterminacy results when current consumption and the real interest rate move in opposite directions. This is true because multiple stationary paths are possible if the paths of the endogenous variables are *not* explosive (that is, $|\hat{c}_{t+1}| < |\hat{c}_t|$) . This guarantees that the real interest rate (which equals consumption growth) will be inversely related to current consumption. In a labor-only economy, this suggests that indeterminacy is possible if and only if the real interest rate ($\hat{i}_t - \hat{\pi}_{t+1}$) and the labor market distortion, the nominal interest rate, ($\hat{i}_t$) move together. By definition, this occurs when $\tau > 1$.

Backward-Looking Interest Rates ($\alpha = 1$)

Now suppose that the central bank conducts policy according to the backward-looking rule. Substituting into (5) the monetary policy rule, $\alpha = 1$ in equation (1), we have

(7) $\hat{\pi}_{t+1} = \tau\hat{\pi}_t$.

The economy is determinate if and only if $\tau > 1$. An aggressive ($\tau > 1$) backward-looking rule pins down the entire inflation sequence, including the current π_t, so that there is both real and nominal determinacy. That is, if the monetary authority responds aggressively to past inflation, the initial price level (and thus the initial money stock) is also pinned down (recall that $\pi_t \equiv \frac{P_t}{P_{t-1}} - 1$ so that pinning down π_t also determines P_t). This contrasts sharply with both an interest rate peg and the forward-looking rule, where there is always nominal indeterminacy.

The intuition for nominal and real determinacy is as follows: Suppose a 1 percent increase in the current price level P_t (and hence π_t). The backward-looking rule implies that next period's nominal rate must rise τ percent. This increase in the future nominal rate (consumption tax) leads to an increase in current consumption, thereby decreasing today's real rate. The policy rule, however, implies that the current nominal rate does not respond to π_t. Hence, the decline in the real rate must lead to an increase in π_{t+1} (to offset the drop in the real rate). This increase will be greater than the initial increase in π_t (see equation [7] with $\tau > 1$). Continuing down this path would be explosive, so it is not a possible equilibrium path.

This confirms McCallum (1981), who argues that the monetary authority could eliminate the nominal indeterminacy associated with interest rate rules by having a nominal anchor. He suggests that this could be achieved by responding to a nominal variable. Our result, however, also shows that merely responding to a nominal variable, such as past

13 The same equation arises in a general money-in-the-utility-function framework containing today's real money balances (Carlstrom and Fuerst [1999]).

14 See Farmer (1999).

inflation, is not enough. To ensure nominal determinacy, the monetary authority must respond aggressively to past inflation ($\tau > 1$).

In appendix 2, we consider the more realistic case of a mixed rule, in which the monetary authority reacts to both past and expected inflation ($\alpha \neq 0$). We show that to avoid nominal indeterminacy in a flexible-price economy (and thus real indeterminacy in a sticky-price economy) the monetary authority must react aggressively to inflation ($\tau > 1$) and most of their response must be from past inflation ($\alpha > 1/2$).

IV. Sunspots and Money Growth

Do changes in inflation correspond to different levels of money? Yes, from the CIA constraint, different values of inflation (and hence current prices) correspond to different levels of money. That is, nominal indeterminacy is equivalent to having more than one money growth process to support a given interest rate objective.[15] For example, adding an independent, identically distributed (i.i.d.) shock to the money growth process does not affect nominal or real interest rates.

Despite the nature of this indeterminacy, the monetary authority is not spinning a roulette wheel to determine money growth. The key is that expected inflation (and hence money growth) can depend on sunspot events whenever the central bank operates off interest rates so that money growth is endogenous. In effect, sunspot events matter in this model economy because the central bank allows them to matter by passively varying the money supply to hit the interest rate objective in (1).

A natural criticism of the previous analysis is that it was conducted in a flexible-price monetary model in which only anticipated inflation had real effects. For example, many would find it peculiar that there can be nominal indeterminacy without any effect on real variables. It is hard to imagine that this extra price volatility would be of no consequence. We argue below that nominal indeterminacy is important because, in the presence of nominal rigidity, it becomes real indeterminacy. Expected inflation and real activity will now be free whenever the initial price level is free.

V. A Sticky-Price Model

Because changes in money feed directly into prices, a flexible-price model implies that i.i.d. shocks to money would have no effect on real variables.[16] From the CIA constraint, $M_t = P_t c_t$, it is obvious that i.i.d. monetary shocks will have real effects in the presence of sticky prices. If prices are fixed, then changes in money must feed into corresponding changes in consumption, as the rest of this section will illustrate more formally. The end result, however, is that nominal indeterminacy is important because it becomes real with sticky prices.

In the simple type of nominal rigidity considered in this section, prices are fixed one period in advance. Because prices are sticky, we must move away from perfect competition, where firms have no pricing power. Imperfect competition implies that prices will no longer be equal to marginal cost; instead, there will be a markup, $1/z$, over marginal cost.[17]

Real wages will no longer be constant and equal to one ($W_t/P_t = f'(L_t) = 1$). Now, workers will be paid less than their marginal productivity

$$\frac{R_t U_L(t)}{U_c(t)} = \frac{W_t}{P_t} = z_t f'(L_t) = z_t < 1,$$

where z_t is marginal cost (see appendix 3 for details). Notice that marginal cost, z_t, acts like a wage tax of $(1 - z_t)$. Written in log-deviation form, (3b) becomes

$$\text{(3c)} \quad \hat{i}_t = \hat{z}_t - \hat{c}_t, \text{ where } z_t = \ln(z_t) - \ln(z_{ss}).$$

Essentially, z_{ss} is a measure of firms' monopoly power. The smaller becomes, the greater is the monopoly power enjoyed by firms. Thus, in contrast to the perfectly competitive example in the previous section, where all firms earned zero profits ($z_{ss} = 1$), with imperfect competition firms will earn profits ($z_{ss} < 1$).

Solving for $\hat{c}_t$ and $\hat{c}_{t+1}$, and substituting them into (2b), yields

$$\text{(5b)} \quad \hat{i}_{t+1} = \hat{\pi}_{t+1} + \hat{z}_{t+1} - \hat{z}_t,$$

where, as before, the "hats" denote log deviations. Notice that when $\hat{z}_t = 0$ ($z_t = z_{ss}$ for all t), equation (5b) collapses to the flexible-price economy of (5a). That is, with flexible prices, $z_t = z_{ss}$, and imperfect competition would have no bearing on the previous analysis. But how is $\hat{z}_t$ determined when prices are sticky?

When a firm sets its prices in advance, it agrees to produce at the level needed to satisfy demand at the fixed price (see the CIA constraint [4]). To increase production, the firm must hire labor,

15 Hence, sunspots can be avoided if the central bank specifies the exact money growth process used to support the desired interest rate target (see, for example, Coleman, Gilles, and Labadie [1996]) or tightly restricts the money growth process (the well known "minimum-state vector solution" of McCallum [1983, 1999]). Both assumptions, however, amount to moving to a money growth operating procedure.

16 The fact that i.i.d. money shocks have no effect on real interest rates is also at the heart of the nominal indeterminancy in the previous section.

17 See appendix 3 or chapter 5 of Walsh (1998).

bidding up the real wage (and thus marginal cost). From the labor equation, this implies that the real wage (and thus marginal cost) will vary with the level of production. Remember that z_t arises because of firms' pricing power. Pricing in advance implies that this pricing power is determined from the prices currently being charged. Therefore $\hat{z}_t$ is free, except to the extent that the CIA constraint (4) must be satisfied. Dividing the CIA constraint by M_{t-1}, log linearizing, and solving for consumption from (3c), we obtain

$$(8)\quad \hat{z}_t = \hat{g}_t + \hat{i}_t \text{ where } \hat{g}_t = \ln\left(\frac{M_t}{M_{t-1}}\right) - g_{ss} \text{ and } g_{ss} = \pi_{ss}.$$

We have used the fact that the existing prices (P_t) and beginning-of-period money stock (M_{t-1}) are predetermined (fixed).

It is important to notice that the firm's marginal cost, or markup, will be determined if and only if both money growth, $\hat{g}_t$, and nominal interest rates, i_t, are determined. Since nominal indeterminacy implies that money growth is not determined, this suggests that real determinacy with sticky prices requires that there be no nominal or real indeterminacy with flexible prices. Once money growth is determined, marginal cost adjusts to ensure that the CIA constraint (8) is satisfied.

A key insight in showing this is that since prices can adjust after one period, $z_{t+j} = z_{ss}$ for all $j \geq 1$; but at time t, z_t need not equal z_{ss} because prices are predetermined.[18] Equilibrium in this economy, therefore, consists of two separate pieces. When prices are sticky,

$$(5c)\quad \hat{z}_t + (\hat{i}_{t+1} - \hat{\pi}_{t+1}) = 0;$$

and when prices are flexible,

$$(5a)\quad \hat{i}_{t+j} - \hat{\pi}_{t+j} = 0 \qquad \text{for all } j \geq 2.$$

Is z_t pinned down? Consider first the flexible-price part (5a). If the flexible-price economy has *both* real and nominal determinacy, then π_{t+1} is uniquely determined. If this occurs, then sticky prices (5c) imply that z_t is determined. But if there is nominal indeterminacy, so that π_{t+1} is not pinned down, then z_t must also be free. The money balances that support this cycle can then be backed out of the CIA constraint (8). Remarkably, the presence of nominal indeterminacy in the flexible-price economy implies that expected inflation (and thus real variables) are free in a sticky-price economy.

To understand this, consider a forward-looking monetary policy rule. Suppose there is a sunspot increase in expected inflation. The monetary policy rule implies that today's funds rate must increase in response. To achieve this, the monetary authority lowers today's money growth, temporarily lowering output (and hence consumption) and thus increasing the real interest-rate. Given this monetary contraction, firms' preset prices will be too high tomorrow. Therefore, the monetary authority must increase tomorrow's money growth to keep the nominal rate in a neutral position. This increases expected inflation today, which means that the initial increase in expected inflation was self fulfilling.

This analysis suggests that since real indeterminacy results whenever expected inflation is not pinned down, a policy in which the monetary authority targets and stabilizes inflation expectations ($\tau = \infty$ and $\alpha = 0$) might be determinate. The advantage of such a policy is that by stabilizing the price level, the central bank stabilizes marginal cost, so that the real economy behaves like a flexible-price model with stabilized prices. The disadvantage is that this policy is also subject to real indeterminacy, as equation (5a) shows. An expected inflation target pins down π_{t+1}, and from (5a) determines i_{t+1}. But i_t is completely free. With expected inflation pegged, this freedom in the nominal rate corresponds to saying that the real rate (and thus real behavior) is subject to sunspot fluctuations. Therefore, an expected inflation target will have real indeterminacy whether prices are flexible or sticky.

VI. Empirical Relevance

At this point, the reader may ask whether this danger is of more than academic interest. Doesn't every central bank look at current economic conditions in determining monetary policy? The answer is, of course, yes, but this observation does not belie our central point. The key issue is whether current conditions influence policy because of what they tell us about the future or because of what they tell us about the past. The sunspot problem arises when central banks use past information to generate forecasts and use these forecasts to drive monetary policy.

How robust is this analysis to more complicated and realistic interest rate policies? Far from being a razor-edge result, the problem that forward-looking rules lead to indeterminacy is extremely robust to the exact formulation of the policy rule and to the modeling environment. Previous research suggested the opposite, arguing that aggressive current and forward-looking rules did not lead to indeterminacy in more complicated sticky-price environments.

18 As noted earlier, to determine whether a model economy is subject to sunspots fluctuations, it is sufficient to examine the perfect-foresight version of the model for arbitrary initial conditions. The initial condition is the predetermined price.

These models, however, ignored the transactions role of money and did not have capital.[19] Carlstrom and Fuerst (2000) demonstrate that correcting either of these omissions leads once again to the results emphasized in this paper.

But if the use of forecasts is so dangerous, why are inflation-targeting countries currently not experiencing any major problems? First, these countries have small, open economies, so the effect from their own monetary policy (and hence the impact of sunspots on domestic activity) would be muted. Because of the counterfactual comparison, it's also impossible to judge whether they are having problems. Yet even if we accept that everything is currently working well in these inflation-targeting countries, we are not dissuaded from our main point. We would like the central bank to follow a monetary strategy that works well under all circumstances. Inflation-forecast targeting rules may work well some of the time, perhaps even most of the time. But they clearly do not work well under a wide range of conditions and therefore should be avoided.

It is also important to note that the severity of the problem today might not indicate its severity tomorrow. In fact, the severity of the problem is likely to increase over time because sunspots involve a coordination problem. That is, the agents in the economy must reach an implicit agreement on what this sunspot event is and base their forecasts on that agreement. This coordination takes time. Unfortunately, sunspots evolve because of the nature of self-fulfilling expectations, particularly if the true causes of inflation are not completely understood. If some variable is thought to cause inflation, the nature of self-fulfilling expectations is such that incorrectly latching on to this variable will help validate the belief that this variable causes inflation.

Suppose that either the public or the monetary authority falsely believes that capacity utilization in and of itself causes future inflation. Even if changes in capacity utilization have no direct impact on expected inflation, they nonetheless initiate a chain of events that cause expected inflation to rise. Over time, the belief in a direct causal connection between the two will become entrenched because inflation typically increases following high capacity-utilization numbers. To borrow a phrase, over time the public may learn to believe in sunspots.[20] This suggests that while sunspot behavior may not arise immediately, it is probably only a matter of time before it does. But there is no conceptual limit on the size and frequency of these sunspots (or consequently on the volatility of inflation and output).

VII. Conclusion

A fundamental contribution of the last three decades of economic research is that private-sector expectations have an enormous influence on the business cycle and on the effect of government policy changes. This paper illustrates a natural corollary: If monetary policy is based on expected inflation, and expected inflation is influenced by monetary policy, then there is real danger that a forward-looking policy will worsen matters by creating the possibility that extra uncertainty is introduced into the economy.

To avoid indeterminacy, the monetary authority must move aggressively against inflation. The key is that it must react primarily to past inflation rather than expected inflation. The basic problem with a proactive agenda is that money growth is endogenous. A backward-looking interest-rate rule may eliminate self-fulfilling expectations by committing the central bank to moving *future* funds rates in response to *today's* price movements. This mitigates the coordination problem because the monetary authority does not move until long after the public has.[21]

It would be a mistake, however, to conclude that central banks should be completely backward looking. Basing policy on the future is of no consequence, or is actually desirable, if monetary policy is firmly grounded in the past. Similarly, looking entirely ahead is no problem in certain highly volatile times. The difficulty arises only when the monetary authority *consistently* bases the bulk of its actions on the future.

To avoid this, the central bank should place the most weight on past movements in the inflation rate (or output). As long as this link between current interest rates and past inflation is aggressive enough, the central bank can eliminate the possibility of self-fulfilling behavior. An immediate implication of this

19 Clarida, Gali, and Gertler (2000) and Kerr and King (1996) are good examples of this line of research. By "ignoring the transactions role of money," we mean that they assume that end-of-period money affected the money-in-the-utility function. This is equivalent to saying that how much money you conserve on transaction costs is determined by how much money you leave the store with versus how much you had entering it. See Carlstrom and Fuerst (1999) for a discussion. Benhabib, Schmitt-Grohe, and Uribe (2001) are also critical of the modeling structure in Clarida, Gali, and Gertler (2000) and in Kerr and King 1996).

20 See Woodford (1990). Carlstrom and Fuerst (2001) also show that sunspots are learnable if the public has rational expectations over anticipated inflation but the central bank must learn about the coefficients of anticipated inflation. If both the central bank and the public are required to learn, learnability is generally much more difficult to achieve.

21 It doesn't completely end the coordination problem because the public's movement is based on their expectation of the monetary authority's future action. This is why a backward-looking rule must also be sufficiently aggressive to eliminate sunspots.

analysis is that inflation targeting over short horizons, which necessarily involves forecasts, is a potentially dangerous policy because it will always be susceptible to sunspots.

Many may wish to reject this conclusion by arguing that theoretical models ignore some aspects of reality. Surely, for example, no central bank uses a rule as simple as the one posited above. But any model must ignore some components of reality. A good model incorporates the salient features of reality and ignores the rest. An examination of the operating procedures of many inflation-targeting central banks leads us to conclude that one salient characteristic of their policy rule is closely approximated by the forward-looking rule that we write down.[22] Theoretical modeling is particularly important here because only theory can shed light on the effects of a monetary policy regime that has not been used before. In the current context, theory has a clear warning. Central banks may increase the volatility of both inflation and real output in their attempt to minimize such volatility through the use of forecasts.

Appendix 1

The household's maximization problem is given by

$$\text{Max} \sum_{t=0}^{\infty} \beta^t U(c_t, T-L_t),$$

$$\text{s.t.} \quad P_t c_t \geq M_{t-1} + X_t + B_{t-1} R_{t-1} - B_t$$

$$M_t = M_{t-1} + X_t + B_{t-1} R_{t-1} - B_t - P_t c_t + P_t f(L_t),$$

where B_t denotes bond holdings (in zero net supply), and monetary injections, X_t, are assumed to be given to households at the beginning of the period. (The remaining notation is as in the text.) Notice that we assume that the household makes the production decision directly. This is without loss of generality. In the sticky-price model (see appendix 3), it is convenient to separate the firm from the household.

Household optimization is defined by the binding cash constraint and the following Euler equations:

(A1) $U_c(t)/P_t = R_t \ \beta U_c(t+1)/P_{t+1}$

(A2) $U_l(t)/P_t = \beta f_L(t) U_c(t+1)/P_{t+1}$.

Substituting (A1) into (A2), we have

(A3) $U_l(t)/U_c(t) = f_L(t)/R_t$.

Given our functional forms

(A4) $\frac{1}{c_t} = \beta \frac{R_t}{(1+\pi_{t+1})} \left(\frac{1}{c_{t+1}}\right)$

(A5) $c_t = \frac{1}{R_t}$.

Taking logarithms and subtracting their long-term (steady-state) values yields (2b) and (3b), where the approximations used are $\ln(R_t) \approx i_t$ and $\ln(1+\pi_t) \approx \pi_t$.

Appendix 2

Plugging (5a) into (1) yields the following equation:

$$F = \tau[\alpha \hat{\pi}_{t-1} + (1-\alpha)\hat{\pi}_{t+1}] - \hat{\pi}_t.$$

The eigenvalues of this equation are given by

$$\lambda_1 = \frac{1 - \left[1 - 4\alpha\tau^2(1-\alpha)\right]^{1/2}}{2\tau(1-\alpha)} \quad \text{and}$$

$$\lambda_2 = \frac{1 + \left[1 - 4\alpha\tau^2(1-\alpha)\right]^{1/2}}{2\tau(1-\alpha)}$$

22 We are not alone in this belief; see, for example, Taylor (1999).

For determinancy, both of these must lie outside the unit circle, which occurs only if the eigenvalues are complex. This leads us to the two following necessary and sufficient conditions for determinancy:

$$[1 - 4\,\alpha\tau^2\,(1-\alpha)] < 0$$

$$\text{and } \frac{1 - [\,1 - 4\,\alpha\tau^2\,(1-\alpha)]}{4\,\tau^2\,(1-\alpha)^2} < 1.$$

Solving these two equations yields the conditions discussed in the text, namely, that the monetary authority must react aggressively to inflation ($\tau > 1$) and that the bulk of their response must be from past inflation ($\alpha > 1/2$).

Appendix 3

In this appendix, we consider a popular model of monetary non-neutrality—a model with sticky prices. We utilize the standard model of imperfect competition in the intermediate goods market,[23] omitting any discussion of household behavior because it is symmetric with appendix 1. The sole exception is that the firm now faces its own decision problem with an objective of maximizing profits which are then paid out to the representative household.

In this economy, final goods are produced in a perfectly competitive industry that utilizes intermediate goods in production. The CES production function is given by

$$Y_t = \{\int_0^1 [y_t(i)^{(\eta-1)/\eta}]\,di\}^{\eta/(\eta-1)}$$

where Y_t denotes the final good and $y_t(i)$ denotes the continuum of intermediate goods, each indexed by $i \in [0,1]$. The implied demand for the intermediate good is thus given by

$$y_t(i) = Y_t\left[\frac{P_t(i)}{P_t}\right]^{-\eta}$$

where $P_t(i)$ is the dollar price of good i, and P_t is the final goods price.

Intermediate goods firm i is a monopolist producer of intermediate good i. (We henceforth omit the firm-specific notation as all firms are symmetric.) The intermediate goods firm is owned by the household and pays its profits out to the household at the end of each period. Because of the cash-in-advance constraint on household consumption, the firm discounts its profits using $\mu_{t+1} \equiv \beta U_c(t+1)/P_{t+1}$, the marginal utility of \$1 in time t+1. Therefore the sticky price is given by the solution to the following maximization problem:

$$P_t^s = \arg\max E_{t-1}\left\{\mu_{t+1}P_tY_t\left(\frac{P_t^s}{P_t}\right)^{-\eta}\left[\left(\frac{P_t^s}{P_t}\right)-z_t\right]\right\},$$

where E_{t-1} is the expectation conditional on time t–1 information. The firm's optimal preset price is thus given by:

$$P_t^s = \left[\left(\frac{\eta}{\eta-1}\right)\frac{E_{t-1}(\mu_{t+1}P_t^{\eta+1}z_tY_t)}{E_{t-1}(\mu_{t+1}P_t^{\eta}Y_t)}\right]$$

In a model without preset prices, this equation would hold at time t, and thus imply that $z_t = z_{ss}$.

As for production, the intermediate firm hires labor from households utilizing the CRS production function from before. Imperfect competition implies that factor payments are distorted. With z_t as marginal cost, we then have $W_t/P_t = z_t\,f'(L_t) = z_t$. Coupling this condition with the household optimization conditions yields the labor market condition utilized in the text.

23 See, for example, Chapter 5 of Walsh (1998).

References

Barro, Robert J., and David B. Gordon. "A Positive Theory of Monetary Policy in a Natural Rate Model," *Journal of Political Economy*, vol. 91, no. 4 (August 1983), pp. 589–610.

Benhabib, Jess, Stephanie Schmitt-Grohe, and Martin Uribe. "Monetary Policy and Multiple Equilibria," *American Economic Review*, vol. 91, no. 1 (March 2001), pp. 167–86.

Carlstrom, Charles T., and Timothy S. Fuerst. "Interest Rate Rules vs. Money Growth Rules: A Welfare Comparison in a Cash-in-Advance Economy," *Journal of Monetary Economics*, vol. 36, no. 2 (November 1995), pp. 247–67.

_____, and _____."Real Indeterminacy in Monetary Models with Nominal Interest Rate Distortions," *Review of Economic Dynamics* (forthcoming).

_____, and _____."Timing and Real Indeterminacy in Monetary Models," *Journal of Monetary Economics*, vol. 47 (2001), pp. 285–98.

_____, and _____."Forward-Looking vs. Backward-Looking Taylor Rules," Federal Reserve Bank of Cleveland, Working Paper no. 00-09, (August 2000).

_____, and _____. "Learning and the Central Bank," Federal Reserve Bank of Cleveland, unpublished manuscript, 2001.

Clarida, Richard, Jordi Gali, and Mark Gertler. "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics*, vol. 115, no. 1 (February 2000), pp. 147–80.

Coleman, Wilbur John II, Christian Gilles, and Pamela A. Labadie. "A Model of the Federal Funds Market," *Economic Theory,* vol. 7, no. 1 (February 1996), pp. 337–57.

Farmer, Roger E.A. *The Macroeconomics of Self-Fulfilling Prophecies*, 2nd ed. Cambridge, Mass.: MIT Press, 1999.

Greenspan, Alan. Discussion of Forrest Capie, Charles A.E. Goodhart, and Norbert Schnadt, "The Development of Central Banking," in Capie et al., *The Future of Central Banking: Proceedings of the Tercentenary Symposium of the Bank of England*. Cambridge: Cambridge University Press, 1994.

Huxford, Julie, and Michael Reddell. "Implementing Monetary Policy in New Zealand." *Reserve Bank Bulletin*, vol. 59, no. 4 (1996), pp. 309–22.

Kerr, William, and Robert G. King. "Limits on Interest Rate Rules in the IS Model," Federal Reserve Bank of Richmond, *Economic Quarterly*, vol. 82, no. 2 (Spring 1996), pp. 47–75.

Kydland, Finn E., and Edward C. Prescott. "Rules Rather than Discretion: The Inconsistency of Optimal Plans," *Journal of Political Economy*, vol. 85, no. 3 (June 1977), pp. 473–91.

McCallum, Bennett T. "Price Level Determinacy with an Interest Rate Policy Rule and Rational Expectations," *Journal of Monetary Economics*, vol. 8, no. 3 (July 1981), pp. 319–29.

_____. "On Non-Uniqueness in Rational Expectations Models: An Attempt at Perspective," *Journal of Monetary Economics,* vol. 11, no. 2 (March 1983), pp. 139–68.

_____. "Issues in the Design of Monetary Policy Rules," in John B. Taylor and Michael Woodford, eds. *Handbook of Macroeconomics*. Amsterdam: Elsevier Science, North-Holland, 1999.

Sargent, Thomas J., and Neil Wallace. "'Rational Expectations,' the Optimal Monetary Instrument, and the Optimal Money Supply Rule," *Journal of Political Economy,* vol. 83, no. 2 (April 1975), pp. 241–54.

Sherwin, Murray. "Inflation Targeting—The New Zealand Experience," in *Proceedings of the Bank of Canada Conference on Price Stability, Inflation Targets, and Monetary Policy*, 1997 <http://www.bankofcanada.ca/en/conference/con97.htm>.

Taylor, John B. "Discretion versus Policy Rules in Practice," *Carnegie–Rochester Conference Series on Public Policy*, vol. 39 (December 1993) pp. 195–214.

_____, ed. *Monetary Policy Rules*. Chicago: University of Chicago Press, 1999.

Walsh, Carl E. *Monetary Theory and Policy*. Cambridge, Mass.: MIT Press, 1998.

Woodford, Michael. "Learning to Believe in Sunspots," *Econometrica*, vol. 58, no. 2 (March 1990), pp. 277–307.