

NBER WORKING PAPER SERIES

ROBUSTLY OPTIMAL MONETARY POLICY  
WITH NEAR-RATIONAL EXPECTATIONS

Michael Woodford

Working Paper 11896  
<http://www.nber.org/papers/w11896>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
December 2005

I would like to thank Ricardo Reis and Tom Sargent for comments, Mauro Roca for research assistance, and the National Science Foundation for research support through a grant to the NBER. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

©2005 by Michael Woodford. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Robustly Optimal Monetary Policy with Near Rational Expectations  
Michael Woodford  
NBER Working Paper No. 11896  
December 2005  
JEL No. D81, D84, E52

### **ABSTRACT**

The paper considers optimal monetary stabilization policy in a forward-looking model, when the central bank recognizes that private-sector expectations need not be precisely model-consistent, and wishes to choose a policy that will be as good as possible in the case of any beliefs that are close enough to model-consistency. The proposed method offers a way of avoiding the assumption that the central bank can count on private-sector expectations coinciding precisely with whatever it plans to do, while at the same time also avoiding the equally unpalatable assumption that the central bank can precisely model private-sector learning and optimize in reliance upon a precise law of motion for expectations.

The main qualitative conclusions of the rational-expectations analysis of optimal policy carry over to the weaker assumption of near-rational expectations. It is found that commitment continues to be important for optimal policy, that the optimal long-run inflation target is unaffected by the degree of potential distortion of beliefs, and that optimal policy is even more history-dependent than if rational expectations are assumed.

Michael Woodford  
Department of Economics  
Columbia University  
1009B I.A.B., mc3308  
New York, NY 10027  
and NBER  
mw2230@columbia.edu

An extensive literature has considered the optimal conduct of monetary policy under the assumption of rational (or model-consistent) expectations. This literature has found that it is quite important to take account of the effects of the systematic (and hence predictable) component of monetary policy on expectations. For example, it is found quite generally that an optimal policy commitment differs from the policy that would be chosen through a sequential optimization procedure with no advance commitment of future policy. It is also found quite generally that optimal policy is *history-dependent* — a function of past conditions that no longer affect the degree to which it would be possible to achieve stabilization aims from the present time onward.<sup>1</sup>

Both of these conclusions, however, depend critically on the idea that an advance commitment of future policy should change people’s expectations at earlier dates. This may lead to the fear that analyses that assume rational expectations (RE) exaggerate the degree to which a policy authority can rely upon private-sector expectations to be shaped by its policy commitments in precisely the way that it expects them to be. What if the relation between what a central bank plans to do and what the public will expect to happen is not quite so predictable? Might both the case for advance commitment of policy and the case for history-dependent policy be considerably weakened under a more skeptical view of the precision with which the public’s expectations can be predicted?

One way of relaxing the assumption of rational expectations is to model agents as forecasting using an econometric model, the coefficients of which they must estimate using data observed prior to some date; sampling error will then result in forecasts that depart somewhat from precise consistency with the analyst’s model.<sup>2</sup> However, selecting a monetary policy rule on the basis of its performance under a specific model of “learning” runs the risk of exaggerating the degree to which the policy analyst can predict and hence exploit the forecasting errors that result from a particular way of extrapolating from past observations. One might even conclude that the optimal policy under learning achieves an outcome better than any possible rational-expectations equilibrium, by inducing systematic forecasting errors of a kind that happen to serve the central bank’s stabilization objectives. But if such a policy were shown to be possible under some model of learning considered to be plausible (or

---

<sup>1</sup>Both points are discussed extensively in Woodford (2003, chap. 7).

<sup>2</sup>Examples of monetary policy analysis under assumptions of this kind about private-sector expectations include Orphanides and Williams (2005a, 2005b) and Gaspar *et al.* (2005).

even consistent with historical data), would it really make sense to conduct policy accordingly, relying on the public to continue making precisely the mistakes that the policy is designed to exploit?

It was exactly this kind of assumption of superior knowledge on the part of the policy analyst that the rational expectations hypothesis was intended to prevent. Yet as just argued, the assumption of RE also implies an extraordinary ability on the part of the policy analyst to predict exactly what the public will be expecting when policy is conducted in a particular way. In this paper, I propose instead an approach to policy analysis that does *not* assume that the central bank can be certain exactly what the public will expect if it chooses to conduct policy in a certain way. Yet neither does it neglect the fact that people are likely to catch on, at least to some extent, to systematic patterns created by policy, in analyzing the effects of alternative policies. In this approach, the policy analyst assumes that private-sector expectations should not be *too different* from what her model would predict under the contemplated policy — people are assumed to have *near-rational expectations* (NRE). But it is recognized that a range of different beliefs would all qualify as NRE. The policymaker is then advised to choose a policy that would not result in too bad an outcome under *any* NRE, *i.e.*, a *robustly* optimal policy given the uncertainty about private-sector expectations.<sup>3</sup>

## 1 Near-Rational Expectations

I can expound the general conception of robust policy that I wish to propose using an abstract two-period policy game. A vector of endogenous variables  $x_t$  is determined in two successive periods ( $t = 0, 1$ ); there are many possible states of the world  $s$  in period 1, and  $x_1$  may depend on  $s$ . The policymaker chooses a vector of controls  $u_t$  in each period; the value of  $u_1$  may be contingent on the state  $s$ . As a result of optimizing behavior by the private sector, in any equilibrium, the endogenous variables  $(x_0, x_1(\cdot))$  must satisfy a system of functional equations

$$F(y_0, y_1(\cdot); \mu) = 0, \tag{1.1}$$

---

<sup>3</sup>The conception of policy robustness here is similar to the one explored in detail in Hansen and Sargent (2005c), though they do not consider the particular source of uncertainty about policy outcomes treated here.

where  $y_t$  is the vector obtained by stacking  $x_t$  and  $u_t$ , and  $\mu$  is the element of  $\mathcal{M}$  (the set of measures over possible values of  $(s, y_1)$ ) that indicates private-sector expectations in the initial period.

The policymaker wishes to choose a policy  $(u_0, u_1(\cdot))$  so as to minimize an expected loss function

$$E[L(y_0, y_1, s)], \quad (1.2)$$

where the expectation  $E[\cdot]$  is with respect to the measure  $\bar{\mu} \in \mathcal{M}$  indicating the policymaker's expectations in the initial period.<sup>4</sup> In the case of any measure  $\pi$  over the possible states of the world  $s$  and any measurable function  $g(\cdot)$ , let  $\nu_{\pi, g}$  denote the element of  $\mathcal{M}$  with marginal distribution  $\pi$  and such that zero probability is assigned to any outcomes in which  $y_1 \neq g(s)$ . Then the policymaker evaluates the objective (1.2) in the case of an equilibrium  $(y_0, y_1(\cdot))$  using the measure

$$\bar{\mu} = \nu_{\bar{\pi}, y_1}, \quad (1.3)$$

where  $\bar{\pi}$  indicates the policymaker's beliefs about the probability of different possible states of nature  $s$ . (In the beliefs of the policymaker,  $\bar{\pi}$  is given independently of the policy chosen, while  $y_1(\cdot)$  and hence  $\bar{\mu}$  will depend on policy.)

In rational-expectations (RE) policy analysis, the analyst assumes that in any equilibrium, the expectations of the private sector will *also* correspond to the measure  $\bar{\mu} = \nu_{\bar{\pi}, y_1}$ . Hence the analyst associates to any policy  $(u_0, u_1(\cdot))$  an equilibrium  $(y_0, y_1(\cdot))$  that satisfies

$$F(y_0, y_1(\cdot); \nu_{\bar{\pi}, y_1}) = 0,$$

and then evaluates (1.2) using the implied measure (1.3).

I shall suppose instead that the analyst recognizes that private agents may *not* have rational expectations, *i.e.*, that beliefs  $\mu \neq \bar{\mu}$  are possible. But I shall suppose that he nonetheless assumes that  $\mu$  is not *too different* from  $\bar{\mu}$ . One reasonable kind of conformity to demand is to assume that private beliefs be *absolutely continuous* with respect to the analyst's beliefs, which means that private agents will agree with

---

<sup>4</sup>The policymaker is here assumed not to entertain doubts about the correctness of her own expectations; thus I am here not concerned with the main kind of uncertainty emphasized by Hansen and Sargent (2005c).

the analyst about which outcomes have zero probability.<sup>5</sup> This requires that private-sector beliefs should be of the form

$$\mu = \nu_{\pi, y_1} \tag{1.4}$$

for *some* measure  $\pi$ , even if  $\pi$  is not necessarily the same as  $\bar{\pi}$ . (In effect, agents are assumed to correctly understand the equilibrium mapping from states of the world to outcomes, even if they do not also correctly assign probabilities to states of the world, as would be required for an RE equilibrium.)

The assumption of absolute continuity also requires that  $\pi$  be absolutely continuous with respect to  $\bar{\pi}$ . A consequence of this is that there must exist a measurable function  $m(\cdot)$ , with the property that  $E[m] = 1$ , such that for any measurable function  $g(\cdot)$  (specifying a random variable at date 1), the expectation  $\hat{E}[g]$  of this random variable under the distorted probability beliefs of the private sector is equal to<sup>6</sup>

$$\hat{E}[g] = E[mg].$$

This representation of the distorted beliefs of the private sector is useful in defining a measure of the distance of the private-sector beliefs  $\pi$  from those of the policy analyst,  $\bar{\pi}$ . As discussed in Hansen and Sargent (2005a, b, c), the *relative entropy*

$$R(\pi, \bar{\pi}) \equiv E[m \log m]$$

is a distance measure with a number of appealing properties.<sup>7</sup> In particular, distorted beliefs  $\pi$  that are not too different from  $\bar{\pi}$  in the sense that  $R(\pi, \bar{\pi})$  is small are ones

---

<sup>5</sup>In the dynamic problem treated in the application to monetary stabilization policy below, I actually assume only that private beliefs be absolutely continuous *over finite time intervals*, as in Hansen *et al.* (2005). This means that I allow for misspecifications that should be detected in the case of a data sample of infinite length, as long as they are not easy to detect using a finite data set. As Hansen *et al.* discuss, this is necessary if one wants the policy analyst to be concerned about possible misspecifications that continue to matter far in the future. Absolute continuity over finite time intervals suffices for the representation of distorted beliefs proposed in this section to continue to apply in the dynamic setting.

<sup>6</sup>The existence of the function  $m(\cdot)$  is guaranteed by the Radon-Nikodym theorem. In the case of a discrete set of states  $s$ ,  $m(s)$  is simply the ratio  $\pi(s)/\bar{\pi}(s)$ . This way of describing distorted beliefs is used, for example, by Hansen and Sargent (2005a, b) and Hansen *et al.* (2005).

<sup>7</sup>For example,  $R(\pi, \bar{\pi})$  is a positive-valued, convex function of  $\pi$ , uniquely minimized (with the value zero) when  $m_{t+1} = 1$  almost surely (the case of RE).

that (according to the beliefs of the analyst) private agents would not be expected to be able to disconfirm by observing the outcome of repeated plays of the game, except in the case of a very large number of repetitions (the number expected to be required being larger the smaller the relative entropy). One might thus view the distorted beliefs  $\pi$  as more plausible the smaller is  $R(\pi, \bar{\pi})$ .

One way to incorporate a concern on the part of the policy analyst for robustness with regard to this type of uncertainty is to suppose that the analyst wishes to choose a policy  $(x_0, x_1(\cdot))$  that is not too bad (does not imply too high a value of (1.2) under any equilibrium (solution to (1.1) associated with private-sector beliefs of the form (1.4) for which the relative entropy is not too large. Thus we might assume that the policy is chosen to minimize

$$\bar{L}(x_0, x(\cdot)) = \max E[L(y_0, y_1, s)], \quad (1.5)$$

where the maximization in (1.5) is over triples  $(y_0, y_1(\cdot), \pi)$  such that (1.1) is satisfied when  $\mu$  is given by (1.4), and such that

$$R(\pi, \bar{\pi}) \leq \bar{R}, \quad (1.6)$$

for some finite bound  $\bar{R} > 0$ . In this case, the concern for robustness would be modeled in a way analogous to the formalization of ambiguity aversion by Gilboa and Schmeidler (1989).

Alternatively, we can model a concern for robustness in a way analogous to the one that is primarily used by Hansen and Sargent (2005c), who follow the lead of the engineering literature on robust control. Instead of supposing that the “worst-case” near-rational expectations (NRE) contemplated by the analyst are those that maximize (1.2) over a set of possible beliefs defined by the constraint (1.6), we may suppose that the worst-case beliefs (and associated equilibrium outcomes) associated with a given policy are the triple  $(\hat{y}_0, \hat{y}_1(\cdot), \hat{\pi})$  that maximize

$$E[L(y_0, y_1, s)] - \theta R(\pi, \bar{\pi}), \quad (1.7)$$

for some penalty coefficient  $\theta > 0$ , over all possible triples  $(y_0, y_1(\cdot), \pi)$  such that (1.1) is satisfied when  $\mu$  is given by (1.4). Here *no* constraint such as (1.6) is imposed on the distorted beliefs that may be considered, but beliefs that are less plausible (from the point of view of the analyst) are more heavily penalized in the objective (1.7).

Thus the analyst will only worry about possible distorted private-sector beliefs that ought to be easy to disconfirm in the case that this particular kind of difference in beliefs would be especially problematic for the particular policy under consideration.<sup>8</sup>

This is the definition of worst-case NRE that I shall use here. The policy analyst is assumed to choose a policy  $(x_0, x_1(\cdot))$  that minimizes  $\hat{L}(x_0, x(\cdot))$ , the maximized value of (1.7) under beliefs (1.3), obtained when  $(y_0, y_1(\cdot))$  are the worst-case NRE beliefs consistent with the policy  $(x_0, x_1(\cdot))$ .<sup>9</sup> One can think of this as the Stackelberg equilibrium of a game between the policymaker and a “malevolent agent” who chooses the private-sector beliefs  $\pi$  that will most embarrass the policymaker.<sup>10</sup> Robust policy in this sense approaches the optimal policy commitment under RE in the limit as  $\theta$  is made unboundedly large, so that the beliefs of the private sector are assumed to be given by  $\pi = \bar{\pi}$  regardless of the policy chosen.

The robust policy problem considered here is related to, though not quite the same as, the type of problem considered by Hansen and Sargent (2005c, chap. 16). Hansen and Sargent discuss a class of “Stackelberg problems” in which a “leader” chooses

---

<sup>8</sup>Maccheroni *et al.* (2004, 2005) show that choosing so as to minimize an objective of the form (1.7) is consistent with a set of axioms for choice under ambiguity aversion only slightly weaker than those of Gilboa and Schmeidler. Both the objective (1.7) and this one are only two members of a broader family that they characterize; the Hansen-Sargent “multiplier preferences” are convenient for my purposes.

<sup>9</sup>Alternatively, one might suppose that the policy analyst is assumed to choose a policy  $(x_0, x_1(\cdot))$  that minimizes  $L^\dagger(x_0, x(\cdot))$ , the value of (1.2) under the same worst-case NRE beliefs. The case assumed in the text corresponds to “variational preferences” of the kind discussed by Maccheroni *et al.* (2004, 2005), and also to the kind of “multiplier robust control problem” treated extensively by Hansen and Sargent (2005c). Apart from the appeal of the axiomatic foundations offered by Maccheroni *et al.* for their representation of preferences, this formulation has the advantage of making the objectives of the policy analyst and the “malevolent agent” perfectly opposed, so that the “policy game” between them is a zero-sum game. This can have advantages when characterizing the solution, though I have not relied on this aspect of the game in the analysis below. The monetary stabilization policy problem is analyzed under the alternative assumption in Woodford (2005), and the same qualitative results are obtained.

<sup>10</sup>Under the assumption made here about the policymaker’s objective, the game is zero-sum, and so under certain regularity conditions (that apply in the application below, for example), the Stackelberg equilibrium is also the Nash equilibrium; one could then analyze a “multiplier game” analogous to the one treated in Hansen and Sargent (2005c, chap. 6). Such a change in the timing of moves by the two “players” is not innocuous, instead, under the alternative objective for the policymaker mentioned in the previous footnote.



a policy taking into account not only the optimizing response of the “follower” to the policy, but also the fact that the follower optimizes under distorted beliefs (*i.e.*, distorted from the point of view of the leader), as a result of the follower’s concern for possible model misspecification.<sup>11</sup> The problem considered here is similar, except that here the policy analyst is worried about the NRE beliefs that would be *worst for her own objectives*, while in the Hansen-Sargent game, the leader anticipates that the follower will act on the basis of the distorted beliefs that would imply *the worst outcome for the follower himself*.<sup>12</sup>

One might think that this difference should not matter in practice, if the policy-maker’s objective coincides with that of the “follower” — as one might think should be the case in an analysis of optimal policy from the standpoint of public welfare. But in the application to stabilization policy below, the private sector is not really a single agent, even though I assume that all price-setters share the same distorted beliefs. It is not clear that allowing for a concern for robustness on the part of individual price-setters would lead to their each optimizing in response to common distorted beliefs, that coincide with those beliefs under which *average* expected utility is lowest.

But more crucially, even in a case where the private sector is made up of identical agents who each solve precisely the same problem, the distorted beliefs that matter in the Hansen-Sargent analysis are those that result in an equilibrium  $(y_0, y_1(\cdot))$  with the highest possible value of  $\hat{E}[L(y_0, y_1, s)]$ , *i.e.*, the greatest *subjective losses* from the point of view of the private sector. In the problem considered here, instead, the NRE beliefs that matter are those that result in an equilibrium with the highest possible value of  $E[L(y_0, y_1, s)]$ ; even if the loss function is identical for the policymaker and the private sector, I assume that it is the *policymaker’s* evaluation of expected losses that matters for robust policy analysis.

In the case that the objective of public policy is assumed to be private welfare, this choice might not be considered obvious; there is always some ambiguity about what

---

<sup>11</sup>Hansen and Sargent also allow for a concern with potential misspecification on the part of the leader, but in the limiting case of their setup in which  $\Theta = \infty$  while  $\theta < \infty$ , only the follower contemplates that the common “approximating model” may be incorrect; the leader regards it as correct, but takes account of the effect on the follower’s behavior of the follower’s concern that the model may be incorrect.

<sup>12</sup>I also consider a different class of possible distorted probability beliefs (Hansen and Sargent allow only for shifts in the mean of the conditional distribution of possible values for the disturbances) and use a different measure of the degree of distortion of PS beliefs (relative entropy).

it should mean for policy to be welfare-maximizing in the case that private agents are regarded by the policy analyst as being mistaken about their situation. Here I take the view that the policy analyst (that in this paper, at least, has no doubts about the correct beliefs) should evaluate private welfare from the point of view of what she believes to be the true likelihood of alternative outcomes. One might also consider the alternative assumption, and define robustly optimal policy as the policy that minimizes  $\hat{E}[L(y_0, y_1, s)]$ . In the application considered next, this alternative assumption would lead to a much more trivial problem: the robustly optimal policy commitment would be exactly like an optimal policy commitment under RE, if the worst-case NRE beliefs were treated as true.<sup>13</sup> Here I consider instead the harder problem of how to choose a robustly optimal policy from the point of view of the policy analyst’s own probability beliefs.

## 2 An Application to Monetary Stabilization Policy

The example considered here weakens the assumption regarding private-sector expectations in the well-known analysis by Clarida *et al.* (1999) of optimal monetary policy in response to “cost-push shocks.” It is assumed that the central bank can bring about any desired state-contingent evolution of inflation  $\pi_t$  and of the output gap  $x_t$  consistent with the aggregate-supply relation

$$\pi_t = \kappa x_t + \beta \hat{E}_t \pi_{t+1} + u_t, \quad (2.1)$$

where  $\kappa > 0, 0 < \beta < 1$ ,  $\hat{E}_t[\cdot]$  denotes the common (distorted) expectations of the private sector (more specifically, of price-setters — I shall call these *PS* expectations) conditional on the state of the world in period  $t$ , and  $u_t$  is an exogenous cost-push

---

<sup>13</sup>This would mean, for example, that the optimal policy commitment could be implemented through commitment to perfectly stabilize a certain linear combination of the log price level and the output gap, as discussed in Woodford (2003, chap. 7). The quantitative form of the optimal target criterion would be completely unaffected by the central bank’s degree of concern for possible forecast error on the part of the private sector. The possibility of NRE beliefs would only have to be taken into account when implementing policy; for example, when evaluating the short-run tradeoff between inflation and the output gap, in order to produce an outcome that satisfies the target criterion.

shock. The analysis is here simplified by assuming that all PS agents have common expectations (though these may not be model-consistent); given this, the usual derivation<sup>14</sup> of (2.1) as a log-linear approximation to an equilibrium relation implied by optimizing price-setting behavior follows just as under the assumption of RE.

The central bank's (CB) policy objective is minimization of a discounted loss function

$$E_0 \sum_{t=0} \beta^t \frac{1}{2} [\pi_t^2 + \lambda(x_t - x^*)^2] \quad (2.2)$$

where  $\lambda > 0$ ,  $x^* \geq 0$ , and the discount factor  $\beta$  is the same as in (2.1). Here  $E_t[\cdot]$  denotes the conditional expectation of a variable under the CB beliefs, which I shall treat as the "true" probabilities, since the analysis is conducted from the point of view of the CB, which wishes to consider the effects of alternative possible policies. I do not allow for any uncertainty on the part of the CB about the probability with which various "objective" states of the world (histories of exogenous disturbances) occur, in order to focus on the issue of uncertainty about PS expectations. The CB believes that the exogenous states  $s_t$  evolve according to a law of motion

$$s_{t+1} = As_t + Bw_{t+1} \quad (2.3)$$

for some matrices  $A, B$ , where the random vector  $w_{t+1}$  is i.i.d. with distribution  $N(0, I)$ ; the cost-push shock each period is then given by  $u_t = v's_t$ . Thus the vector  $s_t$  describes all information available at time  $t$  about current or future "fundamentals". Note that the law of motion (2.3) is *not* assumed to be correctly understood by the PS.

I shall suppose that the central bank chooses (once and for all, at some initial date) a state-contingent policy  $\pi_t = \pi(h_t)$ , where  $h_t \equiv (w_t, w_{t-1}, \dots)$  is the history of realizations of the exogenous disturbances. I assume that commitment of this kind is possible, to the extent that it proves to be desirable; and we shall see that it is desirable to commit in advance to a policy different from the one that would be chosen *ex post*, once any effects of one's decision on prior inflation expectations could be neglected. I also assume that there is no problem for the central bank in *implementing* the state-contingent inflation rate that it has chosen, once a given situation  $h_t$  is reached.<sup>15</sup> This is likely to require that someone in the central bank

---

<sup>14</sup>See, e.g., Woodford (2003, chap. 3).

<sup>15</sup>Even so, the assumption that the central bank commits itself to a state-contingent path for

can observe exactly what PS inflation expectations are at the time of implementation of the policy (in order to determine the nominal interest rate required to bring about a certain rate of inflation); I assume uncertainty about PS expectations only at the time of selection of the state-contingent policy commitment. Note that any such strategy  $\pi(\cdot)$  implies a uniquely defined state-contingent evolution of both inflation and the output gap (given PS beliefs), using equation (2.1), and thus a well-defined value for CB expected losses (2.2).

As in section 1, I shall assume that NRE require that PS beliefs about the economy's evolution over any finite horizon (and in particular, PS beliefs about the probability of various states in the following period) be absolutely continuous with respect to those of the CB. Hence there exists a process  $\{m_{t+1}\}$  with

$$m_{t+1} \geq 0 \quad \text{a.s.}, \quad E_t[m_{t+1}] = 1,$$

such that

$$\hat{E}_t[X_{t+1}] = E_t[m_{t+1}X_{t+1}]$$

for any random variable  $X_{t+1}$ . The degree of distortion of PS beliefs can furthermore be measured by the (discounted) relative entropy

$$E_0 \sum_{t=0}^{\infty} \beta^t m_{t+1} \log m_{t+1},$$

as in Hansen and Sargent (2005a). The presence of the discount factor  $\beta^t$  in this expression means that the CB's concern with potential PS misunderstanding doesn't vanish asymptotically; this makes possible a time-invariant characterization of robustly optimal policy in which the concern for robustness has nontrivial consequences.<sup>16</sup>

Consequently, in the case of any policy commitment  $\{\pi_t\}$  contemplated by the CB, the "worst-case" NRE beliefs considered by the CB are given by the process  $\{m_{t+1}\}$  that maximizes

$$E_0 \sum_{t=0}^{\infty} \beta^t \frac{1}{2} [\pi_t^2 + \lambda(x_t - x^*)^2] - \theta E_0 \sum_{t=0}^{\infty} \beta^t m_{t+1} \log m_{t+1} \quad (2.4)$$

---

inflation, rather than to a Taylor rule or to the satisfaction of some other form of target criterion, is not innocuous. Using this representation of the policy commitment would be innocuous in a RE analysis like that of Clarida *et al.* (1999), since one is effectively choosing from among all possible REE. But here different representations of policy do not lead to the same set of equilibrium allocations being consistent with near-rational expectations.

<sup>16</sup>See Hansen *et al.* (2005) for discussion of this issue, in the context of a continuous-time analysis.

subject to the constraint that  $E_t m_{t+1} = 1$  at all times, where at each date  $x_t$  is implied by the equation

$$\pi_t = \kappa x_t + \beta E_t[m_{t+1} \pi_{t+1}] + u_t. \quad (2.5)$$

Here  $\theta > 0$  is a multiplier that indexes the degree of concern for robustness of policy with respect to non-RE beliefs.

This problem for the “malevolent agent” is in turn equivalent to a sequence of problems in which for each possible history  $h_t$ , a function specifying  $m_{t+1}$  as a function of the realization of  $w_{t+1}$  is chosen so as to maximize

$$\frac{1}{2}[\pi_t^2 + \lambda(x_t - x^*)^2] - \theta E_t[m_{t+1} \log m_{t+1}] \quad (2.6)$$

subject to the constraint that  $E_t m_{t+1} = 1$ , where again  $x_t$  is implied by (2.5). Worst-case NRE then determine the expected output effect of any given state-dependent inflation commitment, according to a time-invariant relation of the form

$$x_t = x^{pess}(u_t, \pi_t, \pi_{t+1}(\cdot)), \quad (2.7)$$

where  $\pi_{t+1}(\cdot)$  specifies  $\pi_{t+1}$  as a measurable function of  $w_{t+1}$ . The degree of distortion of PS beliefs under the worst-case NRE is similarly indicated by a time-invariant function

$$E_t[m_{t+1} \log m_{t+1}] = R^{pess}(u_t, \pi_t, \pi_{t+1}(\cdot)) \quad (2.8)$$

indicating the relative entropy of the worst-case PS beliefs. A robustly optimal policy commitment by the CB is then one that minimizes the maximized value of (2.4), which is to say, that minimizes the objective function obtained by substituting (2.7) for the output gap and (2.8) for the relative-entropy term in (2.4).

This problem can be given a recursive structure if we add an additional constraint, assuming that the initial inflation commitment  $\pi_0(w_0)$  is exogenously given.<sup>17</sup> Let  $J(\pi_0; s_0)$  be the min-max value of (2.4), conditional on a particular initial state. Then under robustly optimal policy, each period the function  $\pi_{t+1}(\cdot)$  is chosen, given the prior inflation commitment  $\pi_t$  and the state  $s_t$ , so as to minimize

$$\begin{aligned} \frac{1}{2}\pi_t^2 + \frac{\lambda}{2}(x^{pess}(u_t, \pi_t, \pi_{t+1}(\cdot)) - x^*)^2 & - \theta R^{pess}(u_t, \pi_t, \pi_{t+1}(\cdot)) \\ & + \beta E[J(\pi_{t+1}(w); As_t + Bw)], \end{aligned} \quad (2.9)$$

---

<sup>17</sup>The same kind of initial commitment defines an optimal policy “from a timeless perspective” in the RE analysis presented in Woodford (2003, chap. 7).

where the expectation is over possible realizations of  $w$ . The minimized value of (2.9) is then precisely the value function  $J(\pi_t; s_t)$ . This constrained version of the robust policy problem is of interest because (as a result of its recursive form) it results in a time-invariant rule for robustly optimal policy.

This recursive structure implies that there is no need for the CB to commit itself more than a period in advance. However, it is important that the state-contingent inflation commitments be chosen at least a period in advance, rather than waiting until the state  $s_{t+1}$  is known and then choosing  $\pi_{t+1}$  to minimize  $J(\pi_{t+1}; s_{t+1})$ . The latter (purely discretionary) approach to policy will not achieve as low a value of (2.9), and hence not as low a value of (2.4) under the worst-case NRE beliefs, as will the approach of choosing a state-contingent inflation commitment each period for the following period. The reason for this advantage of policy commitment is exactly the same, of course, as in the RE analysis of optimal policy in this model (treated in Clarida *et al.*, 1999, and Woodford, 2003, chap. 7).

### 3 Robustly Optimal Linear Policy

Rather than seeking to characterize fully optimal policy in the sense defined above, I shall here characterize the optimal policy within a more restrictive class of *linear policies*. By a linear policy I mean one in which each period's state-contingent inflation commitment is of the form

$$\pi_{t+1}(w_{t+1}) = p_t^0 + p_t^1 w_{t+1}, \quad (3.1)$$

where  $p_t^0$  is some function of  $h_t$  and  $p_t^1$  depends only on  $t$ .<sup>18</sup> The optimal policy commitment under RE is linear in this sense; hence a consideration of this special family of policies suffices to indicate a direction in which it is desirable to change the CB's policy commitment as a result of concern for robustness.

We begin by characterizing the worst-case NRE in the case of an arbitrary linear policy. One notes that an interior solution to the problem of maximizing (2.6) exists

---

<sup>18</sup>It will turn out that in the case of the optimal linear policy,  $p_t^0$  is also a *linear* function of  $h_t$ , but one does not need to impose that.

only if<sup>19</sup>

$$|p_t^1|^2 < \frac{\theta}{\beta^2} \frac{\kappa^2}{\lambda}. \quad (3.2)$$

Otherwise, the objective (2.6) is *convex*, and the worst-case expectations involve extreme distortion, resulting in unbounded losses for the CB. Obviously, it is optimal for the CB to choose a linear policy such that  $p_t^1$  satisfies the bound (3.2) at all times. This provides an immediate contrast with optimal policy under RE, where the optimal vector  $p^1$  (which is constant over time) is proportional to  $\sigma_u$ , the standard deviation of the cost-push shocks.<sup>20</sup> At least for large values of  $\sigma_u$ , it is evident that concern for robustness leads to *less sensitivity* of inflation to cost-push disturbances (smaller  $|p_t^1|$ ). One also observes that it leads to a failure of *certainty equivalence*, as this would require  $|p_t^1|$  to grow in proportion to  $\sigma_u$ .

In the case of a linear policy satisfying (3.2), under the worst-case NRE, the CB fears that the PS will expect  $w_{t+1}$  to be conditionally distributed as  $N(\mu_t, I)$ . If  $p_t^1 = 0$ ,  $\mu_t = 0$ , while if  $p_t^1 \neq 0$ ,

$$\mu_t = (\bar{\pi}_t - p_t^0) \frac{p_t^1}{|p_t^1|^2}, \quad (3.3)$$

where the worst-case inflation expectation (value of  $\hat{E}_t \pi_{t+1}$ ) is given by

$$\bar{\pi}_t = \Delta_t^{-1} \left[ p_t^0 - (\pi_t - u_t - \kappa x^*) \frac{\beta \lambda}{\theta \kappa^2} |p_t^1|^2 \right], \quad (3.4)$$

$$\Delta_t \equiv 1 - \frac{\beta^2}{\theta} \frac{\lambda}{\kappa^2} |p_t^1|^2 > 0. \quad (3.5)$$

The worst-case NRE beliefs distort PS inflation expectations with respect to  $p_t^0$  (the CB's expectation) in the direction opposite to that needed to bring  $x_t$  closer to  $x^*$ ; and this distortion is greater the larger is the sensitivity of (next period's) inflation to unexpected shocks, becoming unboundedly large as the bound (3.2) is approached. As a consequence of this possibility, the CB fears an output gap equal to

$$x_t^{pess} - x^* = \frac{(\pi_t - u_t - \kappa x^*) - \beta p_t^0}{\kappa \Delta_t}. \quad (3.6)$$

---

<sup>19</sup>See the Appendix, section A.1, for derivation of this condition, as well as the results stated in the following two paragraphs. Strictly speaking, it is possible for the inequality (3.2) to be only weakly satisfied, if  $p_t^0$  satisfies a certain linear relation stated in the Appendix; the Appendix treats this case as well, omitted here for simplicity. It is shown in section A.2 that in the robustly optimal linear policy, the inequality is strict.

<sup>20</sup>See, e.g., equation (3.12) below.

Note that  $x_t - x^*$  is larger than it would be under RE by a factor  $\Delta_t^{-1}$ , which exceeds 1 except in the limit in which  $\theta$  is unboundedly large (the RE limit), or if  $p_t^1 = 0$ , so that inflation is perfectly predictable.

The probabilities assigned by the PS to different possible realizations of  $w_{t+1}$  are distorted by a factor  $m_{t+1}$  such that

$$\log m_{t+1} = c_t - \frac{\beta \lambda}{\theta \kappa} (x_t - x^*) \pi_{t+1},$$

where the constant  $c_t$  takes the value necessary in order for  $E_t m_{t+1}$  to equal 1. This implies that the degree of distortion of the worst-case NRE beliefs (as measured by relative entropy) is equal to

$$R_t^{pess} \equiv \hat{E}_t[\log m_{t+1}] = \frac{1}{2} \left[ \frac{\beta \lambda}{\theta \kappa} (x_t - x^*) \right]^2 |p_t^1|^2 \geq 0. \quad (3.7)$$

Note that the degree of distortion against which the policy analyst must guard is greater the larger the degree of inefficiency of the output gap (*i.e.*, the larger is  $|x_t - x^*|$ ), as this increases the marginal cost to the CB's objectives of (maliciously chosen) forecast errors of a given size; and greater the larger the degree to which inflation is sensitive to disturbances (*i.e.*, the larger is  $|p_t^1|$ ), as this increases the scope for misunderstanding of the probability distribution of possible future rates of inflation, for a given degree of discrepancy between CB and PS beliefs (as measured by relative entropy). Of course, it is also greater the smaller is  $\theta$ , the penalty parameter that we use to index the CB's degree of concern for robustness to PS expectational error.

Substituting (3.6) for the output gap and (3.7) for the relative entropy term in (2.4), we obtain a loss function for the CB of the form

$$E_0 \sum_{t=0}^{\infty} \beta^t L(\pi_t; p_t; s_t), \quad (3.8)$$

where  $p_t \equiv (p_t^0, p_t^1)$  and

$$\pi_t = p_{t-1}^0 + p_{t-1}^1 w_t. \quad (3.9)$$

Expression (3.8) indicates the CB's expected losses from a given linear policy  $\{p_t\}$ , under the worst-case NRE beliefs. We wish to minimize this subject to an initial constraint  $\pi_0$ . Moreover, because we do not wish to allow  $p_t^1$  to vary in response to



random shocks, we actually minimize the unconditional expectation of (3.8), integrating over alternative possible initial conditions  $(p_{-1}^0, s_{-1})$  and over alternative possible realizations of  $w_0$ .

A *robustly optimal linear policy* (from a timeless perspective) is then a pair of sequences  $\{p_t^0, p_t^1\}$  that minimize the expected value of (3.8) subject to the law of motion (3.9), given an initial commitment  $p_{-1}^1 = \bar{p}^1$  and integrating over initial conditions  $(p_{-1}^0, s_{-1})$  using a measure  $\rho$ . The value of  $p_t^0$  is allowed to depend on the history  $h_t$ , as well as the particular initial conditions that are drawn from the support of  $\rho$ , but a value for  $p_t^1$  must be chosen that is independent of shock realizations and the same for all initial conditions (which is why the measure  $\rho$  matters). The initial constraints  $(\bar{p}^1, \rho)$  are chosen to be *self-consistent*,<sup>21</sup> which means that under the optimal policy,  $p_t^1 = \bar{p}^1$  for all  $t \geq 0$ , and  $\rho$  is an invariant measure for  $(p_t^0, s_t)$ . One can show that values of  $(\bar{p}^1, \rho)$  exist with this property.

Given  $p_t^1 = \bar{p}^1$ , the loss function  $L(\pi_t; p_t; s_t)$  is a quadratic function of  $(\pi_t, p_t^0, s_t)$ , and the laws of motion (2.3) and (3.9) are linear in these variables. Hence one has a linear-quadratic optimal control problem, and the optimal solution is a linear policy of the form

$$p_t^0 = \mu p_{t-1}^0 + a' s_t + \mu \bar{p}^1 w_t, \quad (3.10)$$

just like the unconstrained optimal policy under RE. A concern for robustness affects the numerical magnitudes of  $\mu$ ,  $a$ , and  $\bar{p}^1$ . But one thing that is not affected is the fact that (3.10) implies stationary fluctuations in the inflation rate around a long-run inflation target of zero. Thus the optimal long-run target is unaffected by the degree of concern for robustness; in particular, allowance for NRE does *not* result in an inflation bias of the kind associated with discretionary policy.<sup>22</sup>

Here I illustrate the quantitative effects of a concern for robustness in an example in which the cost-push shock is purely transitory, so that  $w_t$  is a scalar and  $u_t = \sigma_u w_t$ . Under RE,

$$0 < \mu < 1, \quad a' s_t = -\mu \sigma_u w_t, \quad (3.11)$$

and

$$\bar{p}^1 = \mu \sigma_u. \quad (3.12)$$

---

<sup>21</sup>See Woodford (2003, chap. 7) for the concept of self-consistency invoked here.

<sup>22</sup>On the inflation bias associated with discretionary policy, see Clarida *et al.* (1999) or Woodford (2003, chap. 7).

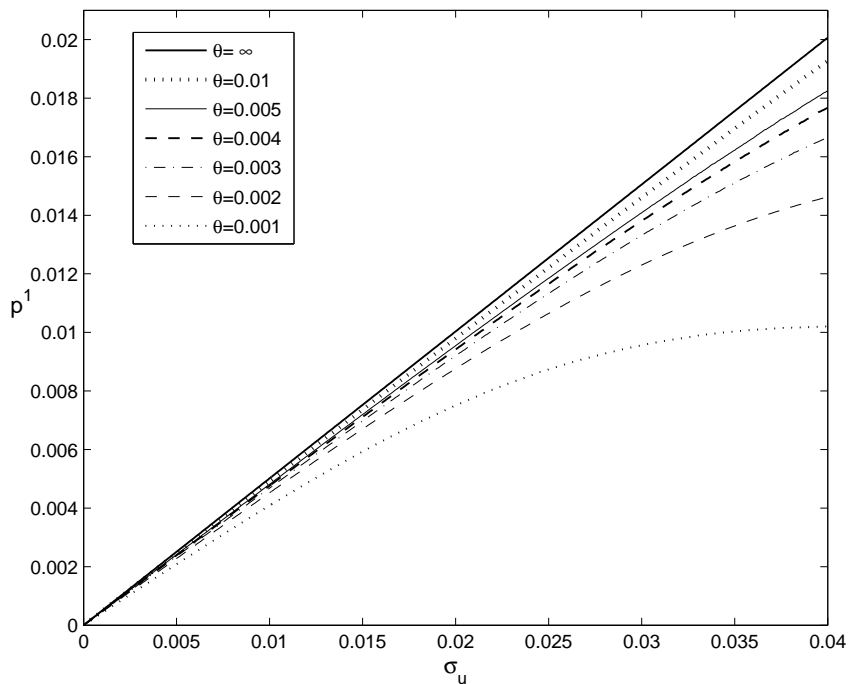


Figure 1: Variation of  $\bar{p}^1$  with  $\sigma_u$ , under alternative degrees of concern for robustness.

With a concern for robustness (finite  $\theta$ ), conditions (3.11) both still hold, but  $\mu$  is now the smaller root of the quadratic equation

$$P(\mu) \equiv \beta\mu^2 - \left(1 + \beta + \frac{\kappa^2\bar{\Delta}}{\lambda}\right)\mu + 1 = 0, \quad (3.13)$$

where  $0 < \bar{\Delta} \leq 1$  is the constant value of (3.5) associated with  $\bar{p}^1$ . It is evident from (3.13) that  $\mu$  is larger the smaller is  $\bar{\Delta}$ ; and since a concern for robustness lowers  $\bar{\Delta}$ , it raises  $\mu$  relative to the RE case. Moreover, contrary to (3.12), one can show that

$$\bar{p}_1 < \mu\sigma_u \quad (3.14)$$

when  $\theta$  is finite.

Figure 1 shows how  $\bar{p}^1$  varies with  $\sigma_u$  for alternative values of  $\theta$ .<sup>23</sup> In the RE case,  $\bar{p}^1$  increases linearly with  $\sigma_u$ , as indicated by (3.12) and as required for certainty-

<sup>23</sup>In this figure, I assume parameter values  $\beta = 0.99$ ,  $\kappa = 0.05$ ,  $\lambda = 0.08$ , and  $x^* = 0.2$ . A low value of  $\lambda$  is justified by the welfare-theoretic foundations of the loss function (2.2) discussed in Woodford (2003, chap. 6).

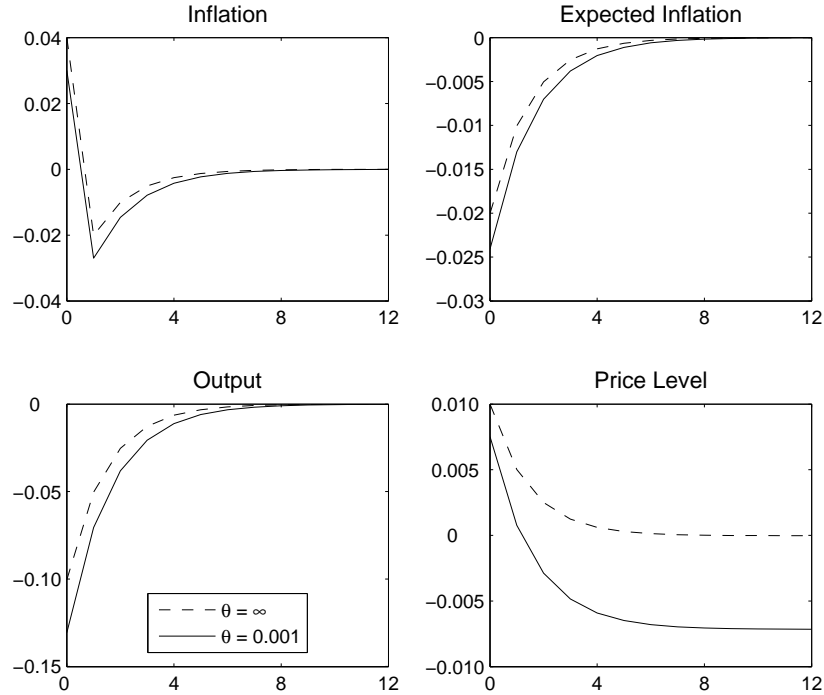


Figure 2: Optimal responses to a positive cost-push shock, with and without concern for robustness.

equivalence. For any given amplitude of cost-push shocks, lower  $\theta$  (greater concern for robustness) results in a lower optimal  $\bar{p}_1$ , indicating less sensitivity of inflation to the current cost-push shock. The extent to which this is true increases in the case of larger shocks; in the case of any finite value of  $\theta$ ,  $\bar{p}_1$  increases less than proportionally with  $\sigma_u$ , indicating a failure of certainty equivalence. In fact,  $\bar{p}_1$  remains bounded above, as required by (3.2).

Thus a concern for robustness results in less willingness to let inflation increase in response to a positive cost-push shock. This is because larger surprise variations in inflation increase the extent to which PS agents may over-forecast inflation, worsening the output/inflation tradeoff facing the CB. This conclusion recalls the one reached by Orphanides and Williams (2005a) on the basis of a model of learning.

At the same time, a concern for robustness increases the degree to which optimal policy is history-dependent. As in the RE case, an optimal commitment involves

a lower inflation rate (on average) in periods *subsequent* to a positive cost-push shock. Moreover, because  $\mu$  is closer to 1 when  $\theta$  is smaller, this effect on average inflation should *last longer*, so that the history-dependence of the optimal inflation commitment is even greater than under RE. And not only should the CB commit to eventually undo any price increases resulting from positive cost-push shocks (as in the RE case); when  $\theta$  is finite, it should commit to eventually reduce the price level *below* the level it would have had in the absence of the shock. This is illustrated in Figure 2 in the case of the numerical example just discussed.<sup>24</sup> The lower right panel shows the impulse response of the log price level; while under rational expectations, the optimal commitment returns the price level eventually to precisely the level that it would have had in the absence of the shock, when  $\theta = 0.001$ , the optimal commitment eventually *reduces* the price level, by an amount about twice as large as the initial price-level increase in response to the shock. The result that the sign of the initial price-level effect is eventually reversed is quite general. Equations (3.9) – (3.11) imply that the cumulative log price increase due to a one-standard-deviation cost-push shock is equal to

$$(\bar{p}^1 - \mu\sigma_u)/(1 - \mu), \quad (3.15)$$

which is zero when (3.12) holds, but negative when  $\theta$  is finite.

Allowance for NRE means that the CB cannot count on its intention to lower inflation (on average) following a positive cost-push shock to lower PS expectations of inflation by as much as the CB's own forecast of future inflation is reduced. (For example, Figure 3 compares the impulse response of PS expected inflation  $\hat{E}_t\pi_{t+1}$  to the response of CB expected inflation  $E_t\pi_{t+1}$ , in the same numerical example as in Figure 2.) But the consequence of this for robustly optimal policy is not that the CB should not bother to try to influence inflation expectations through a history-dependent policy; instead, it is optimal to commit to adjust the subsequent inflation target to an even greater extent and in a more persistent, in order to ensure that inflation expectations are affected even if expectations are not perfectly model-consistent.

---

<sup>24</sup>In the figure, optimal impulse responses to a one-standard-deviation positive cost-push shock are shown, both in the case of infinite  $\theta$  (the standard RE analysis) and for a value  $\theta = 0.001$ . Other parameter values are as in Figure 1; in addition, it is assumed here that  $\sigma_u = 0.02$ .

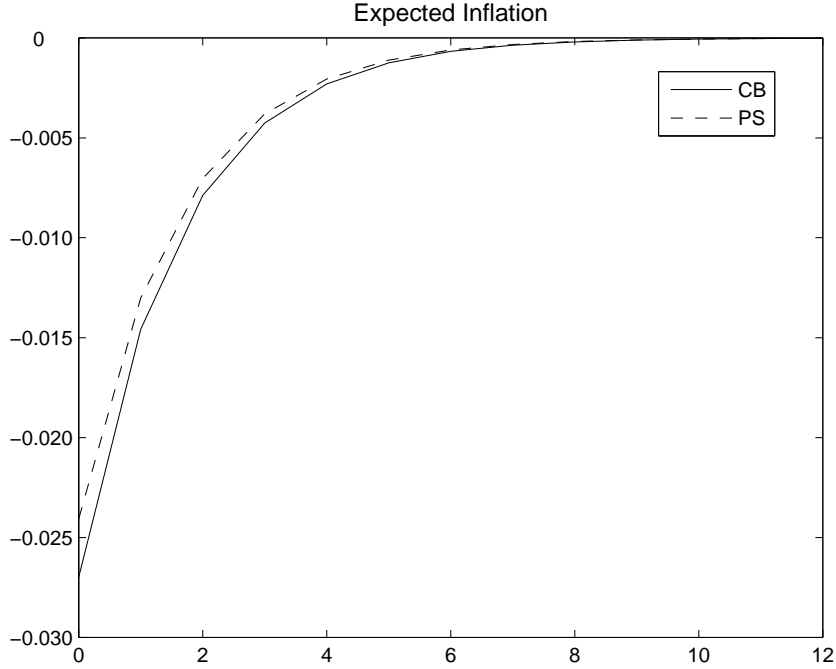


Figure 3: Distortion of PS beliefs, in the worst-case NRE contemplated by the CB when  $\theta = 0.001$ .

## 4 Near-Rational Expectations and the Importance of Policy Commitment

I have observed above that robustly optimal policy involves advance commitment, in a similar way as optimal policy under the assumption of rational expectations. But does the degree to which PS expectations may depart from model-consistency affect the *degree* to which commitment matters? In order to address this question, it is necessary to characterize equilibrium policy under discretionary optimization on the part of a CB that understands that private-sector expectations need not be fully model-consistent, and compare this to the robustly optimal policy under commitment.

Suppose that the objective of the central bank is to minimize (2.4),<sup>25</sup> as above, but that each period the central bank chooses a short-run inflation target  $\pi_t$  after

<sup>25</sup>In appendix A.3, I discuss the way in which my results are modified if one assumes instead that the CB seeks to minimize (2.2), while the malevolent agent seeks to maximize (2.4).

learning the current state  $s_t$ , without making any commitment as to the inflation rate that it may choose at any later dates. Because the payoffs and constraints of both the CB and the malevolent agent in the continuation game at date  $t$  depend on the past only through the state vector  $s_t$ , in a Markov perfect equilibrium (MPE),  $\pi_t$  will depend only on  $s_t$ . I shall assume an equilibrium of this kind;<sup>26</sup> hence there is assumed to exist a time-invariant policy function  $\bar{\pi}(\cdot)$  such that in equilibrium  $\pi_t = \bar{\pi}(s_t)$  each period. Under discretionary optimization, the CB takes for granted the fact that it will choose to follow the rule  $\bar{\pi}(\cdot)$  in all *subsequent* periods, though it is not committed to follow it in the current period. The CB also takes for granted the set of possible NRE beliefs of the PS regarding the economy's future evolution, given that (at least in the view of the CB) the truth is that the exogenous states will evolve according to (2.3), monetary policy will follow the rule  $\bar{\pi}(\cdot)$ , and output will be determined by (2.1). It then chooses an inflation rate  $\pi_t$  to implement in the current period, given its own model of the economy's subsequent evolution and guarding against the worst-case NRE beliefs given that model. In a MPE, the solution to this problem is precisely the inflation rate  $\pi_t = \bar{\pi}(s_t)$ .

I shall formally define a robust MPE as follows. Given a policy rule  $\bar{\pi}(\cdot)$ , let  $V(\pi_0; s_0)$  be the value of the objective (2.4) if the initial state is  $s_0$ , the CB chooses an inflation rate  $\pi_0$  in that initial state and then follows the rule  $\bar{\pi}(\cdot)$  in all periods  $t \geq 1$ , and PS beliefs correspond to the worst-case NRE beliefs given this policy. Then given the inflation rate chosen in any period, the worst-case NRE beliefs  $m_{t+1}(\cdot)$  solve the problem

$$\max_{m_{t+1}(\cdot)} \frac{1}{2} [\pi_t^2 + \lambda(x_t - x^*)^2] - \theta E_t \beta^t m_{t+1} \log m_{t+1} + \beta E_t V(\bar{\pi}(s_{t+1}); s_{t+1}), \quad (4.1)$$

where  $x_t$  satisfies

$$\pi_t = \kappa x_t + \beta E_t [m_{t+1} \bar{\pi}(s_{t+1})] + u_t.$$

A *robust MPE* is then a pair of functions  $\bar{\pi}(\cdot)$  and  $V(\cdot; \cdot)$  such that for any pair  $(\pi_t; s_t)$ ,  $V(\pi_t; s_t)$  is the maximized value of (4.1), and for any state  $s_t$ ,  $\bar{\pi}(s_t)$  is the inflation rate that solves the problem

$$\min_{\pi_t} V(\pi_t; s_t). \quad (4.2)$$

---

<sup>26</sup>The restriction to Markov perfect equilibria is commonplace in the literature on discretionary monetary policy under rational expectations; the equilibrium concept proposed here generalizes the one used by Clarida *et al.* (1999) in their RE analysis of this model.

A *robust linear MPE* is a robust MPE in which  $\bar{\pi}(\cdot)$  is a linear function of the state,

$$\bar{\pi}(s_t) = \pi^* + d' s_t, \quad (4.3)$$

for some long-run average inflation rate  $\pi^*$  and some vector of coefficients  $d$ .

A linear policy (4.3) is an example of the kind of linear policy considered in the previous section, where in terms of the previous notation (3.1) we now have

$$p_t^0 = \pi^* + d' A s_t, \quad p_t^1 = d' B. \quad (4.4)$$

Moreover, because the final term in (4.1) is independent of the choice of  $m_{t+1}(\cdot)$ , the function  $m_{t+1}(\cdot)$  that solves the problem (4.1) is also the one that maximizes (2.6), so that the characterization of worst-case NRE beliefs in appendix A.1 again applies. Once again,  $|p^1|$  must satisfy the bound (3.2) in order for there to be well-defined worst-case beliefs;<sup>27</sup> and when this bound is satisfied, the worst-case beliefs are again described by (3.3) – (3.4).

Given this characterization of worst-case beliefs, the problem (4.2) of the discretionary central bank reduces to

$$\min_{\pi_t} L(\pi_t; \bar{p}(s_t); s_t), \quad (4.5)$$

where  $L(\pi_t; p_t; s_t)$  is the loss function defined in appendix A.2, and  $\bar{p}(s_t)$  is the function defined in (4.4). Since for any  $s_t$ ,  $L$  is a strictly convex, quadratic function of  $\pi_t$ , the discretionary policy  $\bar{\pi}(s_t)$  is just the solution to the first-order condition

$$L_\pi(\pi_t; \bar{p}(s_t); s_t) = 0.$$

This linear equation in  $\pi_t$  is easily solved, yielding

$$\bar{\pi}(s_t) = \frac{\lambda}{\kappa^2 \bar{\Delta} + \lambda} [\kappa x^* + u_t + \beta \bar{p}^0(s_t)]. \quad (4.6)$$

This in turn implies that  $\bar{\pi}(\cdot)$  is indeed a linear function of the form (4.3), where

$$\pi^* = \frac{\lambda \kappa x^*}{\kappa^2 \bar{\Delta} + (1 - \beta) \lambda} > 0, \quad (4.7)$$

---

<sup>27</sup>In the case of discretionary policy, I can no longer argue that the CB will surely choose a policy that satisfies (3.2), in order to avoid unbounded losses. For now the CB is assumed to choose  $\pi_{t+1}$  without taking into account the effect of the way in which the dependence of  $\pi_{t+1}$  on  $w_{t+1}$  affects the worst-case choice of  $m_{t+1}(\cdot)$ , given that the distorted PS beliefs are a historical fact by the time that  $\pi_{t+1}$  is chosen. Nonetheless, there can be no well-defined equilibrium in which (3.2) is violated.

$$d' = \lambda[\kappa^2\bar{\Delta}I + \lambda(I - \beta A)]^{-1}v'. \quad (4.8)$$

In both of these expressions,  $0 < \bar{\Delta} \leq 1$  is the constant value of (3.5) implied by the constant vector  $p^1$  defined in (4.4).

A robust linear MPE exists if we can solve these equations. While equations (4.7) – (4.8) give unique solutions for  $\pi^*$  and  $d$  (and hence for the linear function  $\bar{\pi}(\cdot)$ ) on the assumption that we already know the value of  $p^1$  (and hence the value of  $\bar{\Delta}$ ), the value of  $p^1$  depends on the solution for  $d$ , as indicated in (4.4). There is in fact a fixed-point problem to solve, and depending on parameter values, there may be a unique fixed point, multiple fixed points, or none at all; in the latter case, no robust linear MPE exists.

These possibilities can be illustrated by again considering the case of i.i.d. cost-push shocks, as in section 3. In this case, the fixed-point problem reduces to finding values  $\bar{p}^1, \bar{\Delta}$  that satisfy the two equations

$$\bar{p}^1 = \frac{\lambda}{\kappa^2\bar{\Delta} + \lambda}\sigma_u, \quad (4.9)$$

$$\bar{\Delta} = 1 - \frac{\beta^2}{\theta} \frac{\lambda}{\kappa^2} |\bar{p}^1|^2, \quad (4.10)$$

along with the bound (3.2), so that  $0 < \bar{\Delta} \leq 1$ . One can show that if  $\lambda/\kappa^2 \geq 2$ , there is a unique robust linear MPE if  $\sigma_u < \hat{p}^1$ , while no MPE exist if  $\sigma_u \geq \hat{p}^1$ , where  $\hat{p}^1$  is the upper bound on  $|\bar{p}^1|$  defined in (3.2).<sup>28</sup> If instead  $\lambda/\kappa^2 < 2$ , then there is a unique MPE if  $\sigma_u \leq \hat{p}^1$ , but two distinct MPE if  $\hat{p}^1 < \sigma_u < \sigma_u^*$ , where

$$\sigma_u^* \equiv \frac{2}{3\sqrt{3}} \left[ \frac{\theta}{\beta^2} \left( \frac{\kappa^2 + \lambda}{\lambda} \right)^3 \right]^{1/2}. \quad (4.11)$$

There is again a unique MPE in the special case that  $\sigma_u = \sigma_u^*$ , but there exist no MPE if  $\sigma_u > \sigma_u^*$ .<sup>29</sup>

The possibility of multiple solutions is illustrated numerically in Figure 4. Here the parameter values assumed are as in Figure 1, except that now  $\kappa = 0.15$ ,<sup>30</sup> and I

<sup>28</sup>See appendix A.4 for the proof of this result and the ones stated next, and equation (A.20) in the appendix for the definition of  $\hat{p}^1$ .

<sup>29</sup>Regardless of the value of  $\sigma_u > 0$ , this bound will be violated in the case of small enough  $\theta$ , which is to say, in the case of a large enough concern for robustness on the part of the CB.

<sup>30</sup>A larger value of  $\kappa$  is used in this example in order to illustrate the possibility of multiple solutions, which do not exist under the baseline calibration.



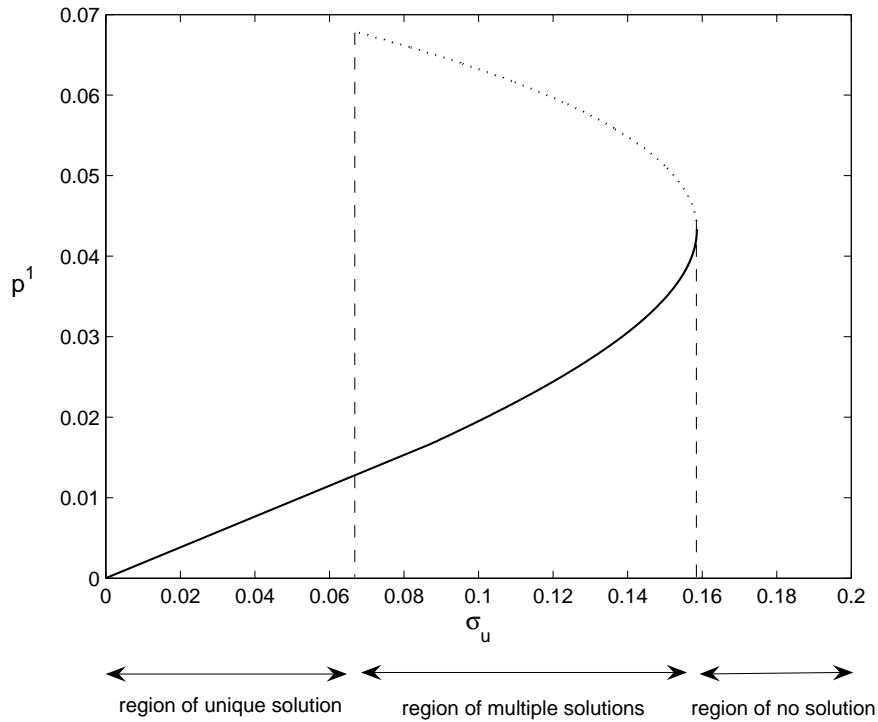


Figure 4: Varying numbers of linear MPE, depending on the size of the cost-push disturbances.

graph the locus of solutions only for the case  $\theta = 0.001$ . A unique solution exists for values of  $\sigma_u$  smaller than 0.068,<sup>31</sup> two solutions exist for values between 0.068 and 0.159, and no solutions exist for larger values of  $\sigma_u$ . In the intermediate range, the second solution (in which inflation is more sensitive to cost-push shocks) is shown by the dotted branch of the locus of fixed points. While these solutions also satisfy the above definition of a robust linear MPE, they are less appealing than the ones on the branch shown as a solid line in the figure, on grounds of what Evans and Honkapohja (2001) refer to as “expectational stability.”

<sup>31</sup>As the graph suggests, there are actually two solutions to the system of equations in this region as well – the dotted branch of the locus of solutions can be extended further to the left. But for values of  $\sigma_u$  this small, the solutions on the dotted branch involve  $\bar{\Delta} < 0$ , and so do not correspond to MPE.

One can reduce the system (4.9) – (4.10) to the single equation

$$\bar{p}^1 = \Phi(\bar{p}^1), \tag{4.12}$$

where  $\Phi(\tilde{p})$  is the value of  $\bar{p}^1$  that satisfies (4.9), when  $\bar{\Delta}$  in this equation is the value obtained by substituting  $\bar{p}^1 = \tilde{p}$  in equation (4.10). Then  $\Phi(p_t^1)$  indicates the degree of sensitivity of inflation to cost-push shocks that would optimally be chosen by a CB choosing under discretion in period  $t$ , if it expects the sensitivity of inflation to cost-push shocks in the following period to be given by  $p_t^1$ .<sup>32</sup> One can show that the lower branch of solutions corresponds to fixed points at which  $0 < \Phi'(\bar{p}^1) < 1$ , while the upper branch corresponds to fixed points at which  $\Phi'(\bar{p}^1) > 1$ . Hence in the former case, an expectation that policy will be near the fixed point far in the future will justify choosing a policy very close to the fixed point now, while in the latter case, even an expectation that policy will be near that fixed point in the distant future will *not* lead the CB to choose policy near that fixed point now — only if future policy is expected to coincide *precisely* with the fixed point will similar behavior be justified now. Hence this fixed point is “unstable” under perturbations of expectations regarding future policy in a way that makes it less plausible that successive central bankers should coordinate on those particular expectations.<sup>33</sup>

What happens in the case of an economy in the region where  $\sigma_u$  is too large for any MPE to exist? (Note that this requires that  $\sigma_u > \hat{p}^1$ .) One observes that  $\Phi(0) > 0$ , and also that  $\Phi(\hat{p}^1) = \sigma_u > \Phi(\hat{p}^1)$ . Then, if there are no fixed points in the interval  $(0, \hat{p}^1)$ ,  $\Phi(p) > p$  over the entire interval.<sup>34</sup> This means that whatever value of  $p^1$  may be expected to describe monetary policy in the following period, a CB that optimizes under discretion will choose a *larger* value in the current period. There is then no Markov perfect equilibrium; but the situation is clearly one in which (an attempt

---

<sup>32</sup>Thus  $\Phi(\cdot)$  is a mapping from the discretionary CB’s “perceived law of motion” to the “actual law of motion” resulting from its optimizing decisions, in the terminology of Evans and Honkapohja (2001).

<sup>33</sup>One can also show that the expectationally stable MPE is an asymptotically stable rest point under adaptive learning dynamics, in which a sequence of central bankers seek to forecast the policies of their successors by extrapolating observed policy in the past, while the expectationally unstable MPE will also be unstable under the learning dynamics. On the connection between expectational stability and stability under adaptive learning dynamics, see generally Evans and Honkapohja (2001).

<sup>34</sup>If instead there are two fixed points, the sign of  $\Phi(p) - p$  changes between them; this is what makes the lower solution expectationally stable while the upper is unstable.

at) discretionary optimization would be expected to lead to very large responses of inflation to cost-push shocks — there would be no reason for the inflation response to remain within any finite bounds!

In the case of rational expectations (the limiting case in which  $\theta^{-1} = 0$ ), there is always a unique solution, given by

$$\bar{p}^1 = \frac{\lambda}{\kappa^2 + \lambda} \sigma_u > 0. \quad (4.13)$$

This is the characterization of policy under discretion given by Clarida *et al.* (1999); the linearity in  $\sigma_u$  again indicates that a principle of certainty equivalence applies. Comparison with (3.12) indicates that under discretionary policy, inflation responds more strongly to a cost-push shock than under the optimal commitment, according to the RE analysis. Moreover, because  $\pi^* > 0$  in the case of discretion, while the long-run average inflation rate is zero under the optimal commitment, discretionary policy is characterized by an “inflationary bias”. These discrepancies between what policy would be like in the best possible RE equilibrium and what it is like in the MPE with discretionary policy indicate the importance of advance commitment to an optimal decision procedure for monetary policy.

How are these familiar results affected by allowing for near-rational expectations? We see from (4.7) that whenever a robust linear MPE exists, it involves a positive average inflation rate  $\pi^*$ ; so again discretionary policy results in an inflationary bias. Moreover, this equation indicates that  $\pi^*$  is a decreasing function of  $\bar{\Delta}$ ; hence the inflationary bias is *increased* by a concern for robustness on the part of the CB (which makes  $\bar{\Delta}$  less than 1). The problem of excessive sensitivity of the inflation rate to cost-push shocks is also increased by a concern for robustness. We observe from (4.9) that

$$\bar{p}^1 > \frac{\lambda}{\kappa^2 + \lambda} \sigma_u \quad (4.14)$$

when  $\bar{\Delta} < 1$ , so that  $\bar{p}^1$  is larger than in the RE case, described by (4.13). One can also show<sup>35</sup> that if we select the lower-sensitivity MPE as “the” prediction of the model when multiple solutions exist, then the solution for  $\bar{p}^1$  is monotonically decreasing in  $\theta$  over the range of values for which a robust linear MPE exists, which means that  $\bar{p}^1$  is higher the greater the concern for robustness.

---

<sup>35</sup>Again see appendix A.4 for the proof.

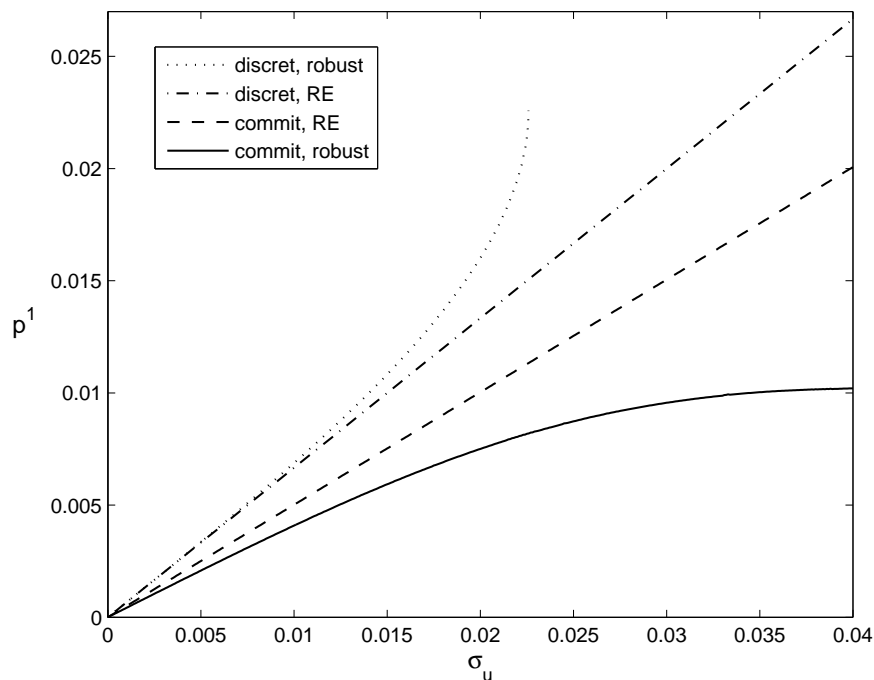


Figure 5: Variation of  $\bar{p}^1$  with  $\sigma_u$ , under discretionary policy and under an optimal commitment, with and without allowance for near-rational expectations.

While a concern for robustness increases the sensitivity of inflation to cost-push shocks under discretionary policy, we found in section 3 that it reduces the sensitivity to cost-push shocks under an optimal commitment. This is illustrated numerically in Figure 5, which extends Figure 1 to show how the equilibrium value of  $\bar{p}^1$  varies with  $\sigma_u$  under discretionary policy as well as under the optimal commitment from a timeless perspective, both with and without an allowance for near-rational expectations.<sup>36</sup> (The two lower curves correspond to cases also shown in Figure 1.) When RE are assumed,  $\bar{p}^1$  is larger under discretionary policy, as just shown; but with a concern for robustness (finite  $\theta$ ), the gap between the values of  $\bar{p}^1$  under discretionary policy and under a robustly optimal linear policy is even larger.

Thus the distortions of policy resulting from optimization under discretion are *increased* when the CB allows for the possibility of near-rational expectations, and

<sup>36</sup>The parameter values used in the figure are again those used in Figure 1. The RE curves assume that  $\theta^{-1} = 0$ , while the ones allowing for NRE beliefs assume that  $\theta^{-1} = 1000$ .

the lessons of the RE analysis become only more important. When the CB's concern for robustness is sufficiently small (*i.e.*,  $\theta$  is large), a robust linear MPE exists, but the degree to which it involves both an excessive average rate of inflation and excessive responsiveness of inflation to cost-push shocks, relative to what would occur under the robustly optimal linear policy, is even greater than is true in the RE analysis. In the case of a sufficiently great concern for robustness, a robust linear MPE fails even to exist; in this case, the dangers of discretionary policy are even more severe, and to an extent much greater than would be suggested by the RE analysis.

## 5 Conclusion

I have shown how it is possible to analyze optimal policy for a central bank that recognizes that private expectations may not be model-consistent, without committing oneself to a *particular* model of expectational error. The approach leads to a one-parameter family robustly-optimal linear policies, indexed by a parameter  $\theta$  that measures the degree of concern for possible misunderstanding of equilibrium dynamics.

Even when the central bank's uncertainty about private expectations is considerable (the case of low  $\theta$ ), calculation of the effects of *anticipations* of the systematic component of policy is still quite an important factor in policy analysis. Optimal policy is still *history-dependent* even when rational expectations are not assumed. Indeed, a concern for robustness only increases the optimal degree of history-dependence.

Moreover, just as in the RE analysis, *commitment* is important for optimal policy. The distortions predicted to result from discretionary policymaking become even more severe when the central bank allows for the possibility of near-rational expectations, so that the importance of commitment is increased. And, as in the RE analysis, a crucial feature of an optimal commitment is a guarantee that inflation will be low and fairly stable. The fact that private beliefs may be distorted does not provide any reason to aim for a higher average rate of inflation, while it does provide a reason for the central bank to resist even more firmly the inflationary consequences of "cost-push" shocks.

# A Appendix: Details of Derivations

## A.1 Worst-Case NRE Beliefs

The problem of the “malevolent agent” in any state of the world at date  $t$  (corresponding to a history  $h_t$  up to that point) is to choose a function specifying  $m_{t+1}$  as a function of the realization of  $w_{t+1}$  so as to maximize

$$\frac{1}{2}[\pi_t^2 + \lambda(x_t - x^*)^2] - \theta E_t[m_{t+1} \log m_{t+1}] \quad (\text{A.1})$$

subject to the constraint that  $E_t m_{t+1} = 1$ , where at each date  $x_t$  is implied by the equation

$$\pi_t = \kappa x_t + \beta E_t[m_{t+1} \pi_{t+1}] + u_t. \quad (\text{A.2})$$

Here I characterize the solution to this problem in the case that the CB follows a linear policy, so that in each possible state at date  $t + 1$  (given the history  $h_t$ ), the inflation rate is given by  $\pi_{t+1} = p_t^0 + p_t^1 w_{t+1}$ , where  $p_t^0$  depends only on  $h_t$  and  $p_t^1$  depends only on the date  $t$ .

It is obvious that the choice of the random variable  $m_{t+1}$  matters only through its consequences for the relative entropy (which affects the objective (A.1)) on the one hand, and its consequences for PS expected inflation (which affects the constraint (A.2) on the other. Hence in the case of any  $\theta > 0$ , the worst-case beliefs will minimize the relative entropy  $E_t[m_{t+1} \log m_{t+1}]$  subject to the constraints that

$$E_t m_{t+1} = 1, \quad E_t[m_{t+1} \pi_{t+1}] = \bar{\pi}_t, \quad (\text{A.3})$$

whatever degree of distortion the PS inflation expectation  $\bar{\pi}_t$  may represent. I first consider this problem. Since  $r(m) \equiv m \log m$  is a strictly convex function of  $m$ , such that  $r'(m) \rightarrow -\infty$  as  $m \rightarrow 0$  and  $r'(m) \rightarrow +\infty$  as  $m \rightarrow +\infty$ , it is evident that there is a unique, interior optimum, such that the first-order condition

$$v'(m_{t+1}) = \phi_{1t} + \phi_{2t} \pi_{t+1}$$

holds in each state at date  $t + 1$ , where  $\phi_{1t}, \phi_{2t}$  are Lagrange multipliers associated with the two constraints (A.3). This implies that

$$\log m_{t+1} = c_t + \phi_{2t} \pi_{t+1} \quad (\text{A.4})$$

in each state, for some constant  $c_t$ . The two constants  $c_t$  and  $\phi_{2t}$  in (A.4) are then the values that satisfy the two constraints (A.3).

Under the assumption of a linear policy,  $\pi_{t+1}$  is conditionally normally distributed, so that (A.4) implies that  $m_{t+1}$  is conditionally log-normal.<sup>37</sup> It follows that

$$\begin{aligned}\log E_t m_{t+1} &= E_t[\log m_{t+1}] + \frac{1}{2}\text{var}_t[\log m_{t+1}] \\ &= c_t + \phi_{2t}p_t^0 + \frac{1}{2}\phi_{2t}^2|p_t^1|^2.\end{aligned}$$

Hence the first constraint (A.3) is satisfied if and only if

$$c_t = -\phi_{2t}p_t^0 - \frac{1}{2}\phi_{2t}^2|p_t^1|^2. \quad (\text{A.5})$$

Under the worst-case beliefs, the PS perceives the conditional probability density for  $w_{t+1}$  to be  $\tilde{f}(w_{t+1}) = m_{t+1}(w_{t+1})f(w_{t+1})$ , where  $f(\cdot)$  is the density for a vector that is distributed as  $N(0, I)$ . It follows from (A.4) and (A.5) that  $\tilde{f}(\cdot)$  is the density for a vector that is distributed as  $N(\mu_t, I)$ , where the bias in the perceived conditional expectation of  $w_{t+1}$  is  $\mu_t = \phi_{2t}p_t^1$ . Hence

$$\hat{E}_t\pi_{t+1} = p_t^0 + p_t^1\mu_t = p_t^0 + \phi_{2t}|p_t^1|^2,$$

and the second constraint (A.3) is satisfied if and only if<sup>38</sup>

$$\phi_{2t} = \frac{\bar{\pi}_t - p_t^0}{|p_t^1|^2}. \quad (\text{A.6})$$

Condition (A.5) then uniquely determines  $c_t$  as well, and  $m_{t+1}$  is completely described by (A.4), once we have determined the value of  $\bar{\pi}_t$  that should be chosen by the “malevolent agent.” Note that the bias  $\mu_t$  is given by expression (3.3), as asserted in the text.

The relative entropy of the worst-case beliefs will then be equal to

$$\begin{aligned}R_t^{pess} = \hat{E}_t[\log m_{t+1}] &= c_t + \phi_{2t}\hat{E}_t\pi_{t+1} \\ &= \frac{1}{2}\frac{(\bar{\pi}_t - p_t^0)^2}{|p_t^1|^2},\end{aligned} \quad (\text{A.7})$$

<sup>37</sup>This is one of the main reasons for the convenience of restricting our attention to linear policies.

<sup>38</sup>Here I assume that  $p_t^1 \neq 0$ . If  $p_t^1 = 0$ , the constraint is satisfied regardless of the distortion chosen by the “malevolent agent,” as long as  $\bar{\pi}_t = p_t^0$ , which is necessarily the case. In this case,  $c_t$  and  $\phi_{2t}$  are not separately identified, but (A.5) suffices to show that  $m_{t+1} = 1$  with certainty.

using (A.5) and (A.6). This is proportional to the squared distance between the PS inflation forecast and that of the central bank; but for any given size of gap between the two, the size of the distortion of probabilities that is required is smaller the larger is  $|p_t^1|$ .<sup>39</sup>

It remains to determine the worst-case choice of  $\bar{\pi}_t$ .<sup>40</sup> It follows from (A.2) that

$$(x_t^{pess} - x^*)^2 = \frac{1}{\kappa^2}(\pi_t - u_t - \kappa x^* - \beta \bar{\pi}_t)^2. \quad (\text{A.8})$$

Substituting this for the squared output gap and (A.7) for the relative entropy in (A.1), we obtain an objective for the “malevolent agent” that is a quadratic function  $Q(\bar{\pi}_t; u_t, \pi_t, p_t)$  of the distorted inflation forecast  $\bar{\pi}_t$ , and otherwise independent of the distorted beliefs; thus  $\bar{\pi}_t$  is chosen to maximize this function. The function is strictly concave (because the coefficient multiplying  $\bar{\pi}_t^2$  is negative) if and only if  $p_t^1$  satisfies inequality (3.2). If the inequality is reversed, the function  $Q$  is instead *convex*, and is minimized rather than maximized at the value of  $\bar{\pi}_t$  that satisfies the first-order condition  $Q_{\bar{\pi}} = 0$ . But in this case, the “malevolent agent” can achieve an unboundedly large positive value of the objective (A.1), as stated in the text; and a robustly optimal policy can never involve a value of  $p_t^1$  this large.

In the case that (3.2) holds with equality,  $Q$  is linear in  $\bar{\pi}_t$ , and it is again possible for the “malevolent agent” to achieve an unboundedly large positive value of the objective through an extreme choice of  $\bar{\pi}_t$ , except in the special case that

$$p_t^0 = \beta^{-1}(\pi_t - u_t - \kappa x^*), \quad (\text{A.9})$$

so that the linear function has a slope of exactly zero. Thus unless  $p_t^0$  satisfies (A.9),  $p_t^1$  must satisfy the bound (3.2) in order for the objective (A.1) to have a finite maximum. Even in the special case that (A.9) holds exactly,  $p_t^1$  must satisfy a variant of (3.2) in which the strict inequality is replaced by a weak inequality.

When (3.2) holds, the maximum value of  $Q$  occurs for the value of  $\bar{\pi}_t$  such that  $Q_{\bar{\pi}} = 0$ . This implies that the worst-case value of  $\bar{\pi}_t$  is the one given by (3.4) – (3.5) in the text. Substituting this solution into (A.7) and (A.8), one obtains the implied output gap (3.6) and and relative entropy (3.7) under the worst-case NRE beliefs,

---

<sup>39</sup>Equation (A.7) again assumes that  $p_t^1 \neq 0$ . In the event that  $p_t^1 = 0$ , it follows from the previous footnote that the relative entropy of the worst-case beliefs will equal zero.

<sup>40</sup>The analysis here assumes that  $p_t^1 \neq 0$ . If  $p_t^1 = 0$ , there is no choice about the value of  $\bar{\pi}_t$ ; it must equal  $p_t^0$ .



as stated in the text. Substituting these expressions into the objective (A.1), one obtains an objective for the CB of the form (3.8), in which the period loss is given by

$$L(\pi_t; p_t; s_t) = \frac{1}{2}\pi_t^2 + \frac{\lambda}{2\kappa^2\Delta_t}[\pi_t - u_t - \kappa x^* - \beta p_t^0]^2, \quad (\text{A.10})$$

where  $\Delta_t$  is the function of  $p_t^1$  defined by (3.5).

When, instead, (3.2) holds with equality, and (A.9) holds as well, the worst-case value of  $\bar{\pi}_t$  is indeterminate, but the maximized value of (A.1) is nonetheless well-defined, and equal to zero. In this case, the period loss function is equal to

$$L(\pi_t; p_t; s_t) = \frac{1}{2}\pi_t^2.$$

When neither this case nor the one discussed in the previous paragraph applies, we can define  $L(\pi_t; p_t; s_t)$  as being equal to  $+\infty$ . The function is then defined (but possibly equal to  $+\infty$ ) for all possible values of its arguments.

Note also that  $L(\pi_t; p_t; s_t)$  is necessarily non-negative, since for any values of the arguments, it is possible for the “malevolent agent” to obtain a non-negative value of (A.1) by choosing  $m_{t+1} = 1$  in all states; the maximized value of (A.1) is then necessarily at least this high. It follows that both the conditional expectations and the infinite sum in (3.8) are sums (or integrals) of non-negative quantities; hence both are well-defined (though possibly equal to  $+\infty$ ) for all possible values of the arguments. Thus the CB objective (3.8) is well-defined for arbitrary state-contingent sequences  $\{p_t\}$  and an arbitrary initial condition  $(\pi_0, s_0)$ .

## A.2 Robustly Optimal Linear Policy

Given the worst-case PS beliefs characterized in the previous section, the problem of the CB is to choose a sequence  $\{p_t\}$  for all  $t \geq 0$  so as to minimize

$$E_\rho E_0 \sum_{t=0}^{\infty} \beta^t L(\pi_t; p_t; s_t), \quad (\text{A.11})$$

where

$$\pi_{t+1} = p_t^0 + p_t^1 w_{t+1} \quad (\text{A.12})$$

and

$$s_{t+1} = A s_t + B w_{t+1} \quad (\text{A.13})$$

for each  $t \geq 0$ , and  $(\pi_0, s_0)$  are given as initial conditions. Here  $E_\rho$  indicates an integral over alternative possible initial conditions  $(p_{-1}^0, s_{-1}, w_0)$  using a measure  $\rho$ , the choice of which is explained in the next section; and it is assumed that  $p_{-1}^1 = \bar{p}^1$ , where the choice of  $\bar{p}^1$  (a single value) is also explained in the next section. I use the notation  $E_t[\cdot]$  to indicate an expectation conditional upon the history  $h_t$ , by which I mean the particular initial conditions  $(p_{-1}^0, s_{-1}, w_0)$  that have been drawn, together with the subsequent realizations of the exogenous disturbances  $(w_1, \dots, w_t)$ . I furthermore suppose that the CB's choice of  $p_t^1$  must depend only on the date  $t$ , while the choice of  $p_t^0$  may depend on the history  $h_t$ .

One can show that the objective (A.11) is a convex function of the sequence  $\{p_t\}$ . I begin by noting that (A.1) is a convex function of  $\pi_t$  and  $x_t$ , for any choice of  $m_{t+1}(\cdot)$ . Then since (A.2) is a linear relation among  $\pi_t, x_t$ , and  $\pi_{t+1}(\cdot)$ , it follows that, taking as given the choice of  $m_{t+1}(\cdot)$ , the value of (A.1) implied by any choice of  $\pi_{t+1}(\cdot)$  by the CB is a convex function of  $\pi_t$  and  $\pi_{t+1}(\cdot)$ . Similarly, since (A.12) is linear, the value of (A.1) implied by any choice of  $p_t$  is a convex function of  $\pi_t$  and  $p_t$ , for any choice of  $m_{t+1}(\cdot)$ . Then since the maximum of a set of convex functions is a convex function, it follows that the maximized value of (A.1) is also a convex function of  $\pi_t$  and  $p_t$ . Thus  $L(\pi_t; p_t; s_t)$  is a convex function of  $(\pi_t, p_t)$ . Finally, a sum of convex functions is convex; this, together with the linearity of (A.12), implies that (A.11) is a convex function of the sequence  $\{p_t\}$ .

Convexity implies that the CB's optimal policy can be characterized by a system of first-order conditions, according to which

$$L_0(\pi_t; p_t; s_t) + \beta E_t L_\pi(\pi_{t+1}; p_{t+1}; s_{t+1}) = 0 \quad (\text{A.14})$$

for each possible history  $h_t$  at any date  $t \geq 0$ , and

$$E_\rho E_0 [L_1(\pi_t; p_t; s_t) + \beta L_\pi(\pi_{t+1}; p_{t+1}; s_{t+1}) w_{t+1}] = 0 \quad (\text{A.15})$$

for each date  $t \geq 0$ . Here  $L_\pi$  denotes  $\partial L / \partial \pi$ ,  $L_0$  denotes  $\partial L / \partial p^0$ , and  $L_1$  denotes  $\partial L / \partial p^1$ . Condition (A.14) is the first-order condition for the optimal choice of  $p_t^0$ , and (A.15) is the corresponding condition for the optimal choice of  $p_t^1$ . The latter condition is required to hold only in its *ex ante* (or unconditional) expected value, because I have defined a linear policy as one under which  $p_t^1$  does not depend on the history of realization of the shocks.

Note that it follows from the characterization in the previous section that for any plan satisfying (3.2), the partial derivatives just referred to are well-defined, and equal to

$$\begin{aligned} L_\pi(\pi_t; p_t; s_t) &= \pi_t + \frac{\lambda}{\kappa^2} \frac{\pi_t - u_t - \kappa x^* - \beta p_t^0}{\Delta_t}, \\ L_0(\pi_t; p_t; s_t) &= -\beta \frac{\lambda}{\kappa^2} \frac{\pi_t - u_t - \kappa x^* - \beta p_t^0}{\Delta_t}, \\ L_1(\pi_t; p_t; s_t) &= \frac{\beta^2}{\theta} \left( \frac{\lambda}{\kappa^2} \right)^2 \left( \frac{\pi_t - u_t - \kappa x^* - \beta p_t^0}{\Delta_t} \right)^2 p_t^1. \end{aligned}$$

Substituting (A.12) for  $\pi_t$  and (3.5) for  $\Delta_t$  in these expressions, one obtains the first-order conditions (A.14) – (A.15) as restrictions upon the sequence  $\{p_t\}$ .

As explained in the text, I wish to find a sequence of functions  $\{\varphi_t(\cdot)\}$  and a value  $\bar{p}^1$  such that the linear policy under which

$$p_t^0 = \varphi_t(h_t), \quad p_t^1 = \bar{p}^1$$

for all  $t \geq 0$  satisfies the first-order conditions (A.14) – (A.15), when the initial measure  $\rho$  is the ergodic measure for the variables  $(p_t^0, s_t, w_{t+1})$  under the policy just specified, and in addition  $p_{-1}^1 = \bar{p}^1$ . I first show that there exists a state-contingent evolution for  $\{p_t^0\}$  that satisfies (A.14) in the case of an arbitrary constant value  $\bar{p}^1$  that satisfies the bound (3.2), and for which there exists a well-defined ergodic measure. Using the ergodic measure  $\rho$  corresponding to a given value of  $\bar{p}^1$ , I then determine the nonlinear equation that  $\bar{p}^1$  must satisfy in order for (A.15) to hold each period under the conjectured solution. Demonstration that a robustly optimal linear policy exists then requires only that one show that there exists a solution  $\bar{p}^1$  to this equation that also satisfies the bound (3.2).

Under the assumption that  $p_t^1 = \bar{p}^1$  for all  $t \geq -1$ , (A.14) is a stochastic linear difference equation for the process  $\{p_t^0\}$  of the form

$$E_t[A(L)p_{t+1}^0] = v_t, \tag{A.16}$$

where

$$\begin{aligned} A(L) &\equiv \beta - \left( 1 + \beta + \frac{\kappa^2 \bar{\Delta}}{\lambda} \right) L + L^2, \\ v_t &\equiv u_t - E_t u_{t+1} - \bar{p}^1 w_t. \end{aligned}$$

(Here  $\bar{\Delta}$  is the constant value of  $\Delta_t$  implied by the constant value  $\bar{p}^1$ .) By factoring the lag polynomial in (A.16), one can easily show that (A.16) has a unique stationary solution, given by

$$p_t^0 = \mu p_{t-1}^0 - \mu E_t[(1 - \beta\mu L^{-1})^{-1} v_t], \quad (\text{A.17})$$

where  $0 < \mu < 1$  is the smaller root of the characteristic equation (3.13) given in the text. Note that a stationary solution exists regardless of the value assumed for  $\bar{p}^1$ . It is then straightforward to solve for the ergodic measure  $\rho$  associated with a given value of  $\bar{p}^1$ .

Equation (A.17) is a solution for the dynamics of  $\{p_t^0\}$  of the kind indicated by equation (3.10) in the text. In the special case in which  $w_t$  is a scalar and  $u_t = \sigma_u w_t$ ,  $v_t = (\sigma_u - \bar{p}^1)w_t$ , and (A.17) reduces to

$$p_t^0 = \mu p_{t-1}^0 - \mu(\sigma_u - \bar{p}^1)w_t. \quad (\text{A.18})$$

Thus we have established conditions (3.11) given in the text. As noted in the text, it is evident from (3.13) that  $\mu$  is monotonically decreasing in  $\bar{\Delta}$ . Since a concern for robustness results in  $\bar{\Delta} < 1$ , while  $\bar{\Delta} = 1$  in the case of rational expectations, we see that a concern for robustness results in a value of  $\mu$  that is larger (closer to 1), implying more persistence in the fluctuations in  $\{p_t^0\}$ .

It remains to determine when condition (A.15) is also satisfied. I first observe that

$$\begin{aligned} E_\rho E_0[L_1(\pi_t; p_t; s_t)] &= \frac{\beta^2}{\theta} \left( \frac{\lambda}{\kappa^2} \right)^2 \frac{\bar{p}^1}{\bar{\Delta}^2} E[(\pi_t - u_t - \kappa x^* - \beta p_t^0)^2] \\ &= \frac{\beta^2}{\theta} \left( \frac{\lambda}{\kappa^2} \right)^2 \frac{\bar{p}^1}{\bar{\Delta}^2} [a + 2b\bar{p}^1 + (\bar{p}^1)^2], \end{aligned}$$

where

$$\begin{aligned} a &\equiv E[(p_{t-1}^0 - u_t - \kappa x^* - \beta p_t^0)^2], \\ b &\equiv E[w_t(p_t^0 - u_t - \kappa x^* - \beta p_t^0)]. \end{aligned}$$

Here  $E[\cdot]$  denotes the expectation under the ergodic measure associated with the dynamics for  $\{p_t^0\}$  indicated by (A.17) — which measure is uniquely defined in the case of a given value of  $\bar{p}^1$ .

Similarly, one can show that

$$\begin{aligned} E_\rho E_0[L_\pi(\pi_{t+1}; p_{t+1}; s_{t+1})w_{t+1}] &= E[\pi_{t+1}w_{t+1}] + \frac{\lambda}{\kappa^2 \bar{\Delta}} E[(\pi_{t+1} - u_{t+1} - \kappa x^* - \beta p_{t+1}^0)w_{t+1}] \\ &= \bar{p}^1 + \frac{\lambda}{\kappa^2 \bar{\Delta}} [\bar{p}^1 + b]. \end{aligned}$$

Hence condition (A.15) is equivalent to

$$f(\bar{p}^1) \equiv \frac{\beta^2}{\theta} \left( \frac{\lambda}{\kappa^2} \right)^2 \frac{c}{\bar{\Delta}^2} \bar{p}^1 + \bar{p}^1 + \frac{\lambda}{\kappa^2 \bar{\Delta}} [\bar{p}^1 + b] = 0, \quad (\text{A.19})$$

where

$$c \equiv a + 2b\bar{p}^1 + (\bar{p}^1)^2.$$

A robustly optimal linear policy then exists if and only if (A.19) has a solution  $\bar{p}^1$  that satisfies the bound (3.2). Of course, in defining the function  $f(\cdot)$ , one must take account of the dependence of  $c$  and  $\bar{\Delta}$  on the value of  $\bar{p}^1$ .

When  $\{p_t^0\}$  evolves in accordance with the stationary dynamics (A.18), the above definitions imply that

$$\begin{aligned} a &= (\kappa x^*)^2 + E\{[(1 - \beta\mu)p_{t-1}^0 - (\sigma_u - \beta\mu(\sigma_u - \bar{p}^1))w_t]^2\} \\ &= (\kappa x^*)^2 + \frac{(1 - \beta\mu)^2 \mu^2}{1 - \mu^2} (\sigma_u - \bar{p}^1)^2 + [(1 - \beta\mu)\sigma_u + \beta\mu\bar{p}^1]^2, \end{aligned}$$

$$\begin{aligned} b &= -\sigma_u - \beta E[p_t^0 w_t] \\ &= -(1 - \beta\mu)\sigma_u - \beta\mu\bar{p}^1. \end{aligned}$$

I furthermore observe that  $a = a_0 + b^2$ , where

$$a_0 \equiv (\kappa x^*)^2 + \frac{(1 - \beta\mu)^2 \mu^2}{1 - \mu^2} (\sigma_u - \bar{p}^1)^2 > 0.$$

Hence

$$c = a_0 + (b + \bar{p}^1)^2 > 0$$

can be signed for all admissible values of  $\bar{p}^1$ . Substituting this function of  $\bar{p}^1$  for  $c$  and (3.5) for  $\bar{\Delta}$  in (A.19) yields a nonlinear equation in  $\bar{p}^1$ , that is solved numerically in order to produce Figure 1.

One can easily show that a solution to this equation in the admissible range must exist. Note first that (3.2) can alternatively be written in the form

$$|\bar{p}^1| < \hat{p}^1 \equiv \frac{\kappa}{\lambda^{1/2}} \frac{\theta^{1/2}}{\beta}. \quad (\text{A.20})$$

I next observe that

$$f(0) = \frac{\lambda}{\kappa^2 \bar{\Delta}} b = -\frac{\lambda}{\kappa^2} (1 - \beta\mu)\sigma_u < 0.$$

On the other hand, in the case of any finite  $\theta$ , as  $p^1 \rightarrow \hat{p}^1$ , the first term in the expression (A.19) becomes larger than the other two terms, so that  $f(p^1) > 0$  for any value of  $p^1$  close enough to (while still below) the bound. Since the function  $f(\cdot)$  is well-defined and continuous on the entire interval  $[0, \hat{p}^1)$ , there must be an intermediate value  $0 < \bar{p}^1 < \hat{p}^1$  at which  $f(\bar{p}^1) = 0$ . Such a value satisfies both (3.2) and (A.19), and so describes a robustly optimal linear policy.

It remains to establish (3.12) and (3.14). When evaluated at the value  $p^1 = \mu\sigma_u$ , the second two terms in (A.19) are equal to

$$-\frac{\lambda}{\kappa^2 \bar{\Delta}} P(\mu)\sigma_u = 0,$$

where  $P(\mu)$  is the polynomial defined in (3.13). Moreover, in the limiting case in which  $\theta \rightarrow \infty$  (the RE case), the first term in condition (A.19) is identically zero, so that  $f(\mu\sigma_u) = 0$ , and  $\bar{p}^1 = \mu\sigma_u$  is a solution.<sup>41</sup> Instead, when  $\theta$  is finite, the first term is necessarily positive, so that  $f(\mu\sigma_u) > 0$ . If  $\mu\sigma_u < \hat{p}^1$ , this implies that there exists a solution to (A.15) such that  $0 < \bar{p}^1 < \mu\sigma_u$ , as asserted in (3.14). If instead  $\hat{p}^1 \leq \mu\sigma_u$ , then (3.14) follows from the result in the previous paragraph. Hence in either case, the robustly optimal policy satisfies (3.14) for any finite  $\theta$ .

### A.3 An Alternative Objective for the Robust Policymaker

In the analysis presented above, it is assumed that the robust policymaker seeks to minimize precisely the same objective (2.4) that the “malevolent agent” seek to maximize through its choice of distorted beliefs for the private sector. This corresponds to the kind of robust decision problem that Hansen and Sargent (2005, chap. 6) call a “multiplier robust control problem,” and to the kind of generalized ambiguity aversion characterized by Maccheroni *et al.* (2004, 2005). Here, instead, I show how the results would differ in the case of the alternative objective assumed in Woodford (2005).

One might alternatively suppose that the policymaker ranks alternative equilibria according to the value of the original loss function (2.2), even though the malevolent agent is expected to prefer the highest possible value of (2.4). The reasoning would be

---

<sup>41</sup>It is easily seen to be the unique solution, since  $f(p)$  is linear in this case. One can also show that this is the optimal policy without restricting attention to linear policies, as is done here; see Clarida *et al.* (1999) or Woodford (2003, chap. 7).

the following. The policymaker is worried about possible outcomes (due to distorted beliefs) that are especially bad in terms of his objective (2.2), but less worried if the distortion of beliefs required is one that should be relatively easily disconfirmed by observation. The worst-case outcome that is feared, in the case of a given policy, is therefore the one that maximizes (2.4). However, when the policymaker considers which policy is least unfortunate under these worst-case beliefs, he ranks the worst-case equilibria associated with alternative policies according to the value of (2.2) in each. The policymaker is concerned by the possibility of an equilibrium with large departures of inflation and/or the output gap from their optimal values, and does not regard it as a mitigating factor that such an equilibrium might involve large expectational errors on the part of the private sector. Ranking outcomes by the value of (2.4) suggests instead that the policymaker is not bothered by failure to stabilize inflation and the output gap, as long as the failure depends on a sufficiently large distortion of PS beliefs.

This modification of the assumed objective of robust policy has the consequence that the “game” between the policymaker and the malevolent agent is no longer zero-sum, which prevents us from using a number of devices that are frequently employed in Hansen and Sargent (2005); for example, we cannot characterize optimal policy by the first-order conditions for an optimal policy commitment in the case of the particular distorted beliefs  $\{m_{t+1}\}$  that are chosen by the malevolent agent.<sup>42</sup> But the method used above to characterize robustly optimal policy assumes a “game” in which the CB is a Stackelberg “leader” and takes into account the optimal response of the malevolent agent, and as a result it is not hard to explain how the calculations would be modified in the case of the alternative objective.

The characterization of worst-case NRE beliefs in section A.1 remains applicable, since the malevolent agent again seeks to maximize (2.4). However, rather than substituting the above solutions for  $m_{t+1}(w_{t+1})$  and  $x_t$  into (A.1), one substitutes them instead into a period loss function that omits the final (relative-entropy) term. Again one obtains an objective for the CB of the form (3.8), but now the period loss

---

<sup>42</sup>Thus a characterization of robustly optimal policy by a coupled system of Euler equations for the decisionmaker and the malevolent agent respectively, as in “game 1” of Hansen and Sargent (2005, chap. 6), is not possible in this case, though it would be possible for the robust policy problem stated in the text.

is given by

$$L(\pi_t; p_t; s_t) = \frac{1}{2}\pi_t^2 + \frac{\lambda}{2\kappa^2\Delta_t^2}[\pi_t - u_t - \kappa x^* - \beta p_t^0]^2, \quad (\text{A.21})$$

rather than (A.10). (Here  $\Delta_t$  is again the function of  $p_t^1$  defined by (3.5).) Note that the *qualitative* conclusion about how the CB's objective is affected by a concern for robustness remains the same as before: the relative weight on the second term in (A.21) is increased the smaller is  $\theta$ , because (for any given linear policy in which  $p_t^1 \neq 0$ )  $\Delta_t$  becomes a smaller positive quantity as  $\theta$  is reduced. The fact that  $\Delta_t$  is replaced by  $\Delta_t^2$  in (A.21) only means that the effect is even stronger in the case of this version of the robust policy problem. (Under the present objective, the worse outcome for the output gap is no longer partially mitigated by the fact that PS beliefs are also more distorted in the case of lower  $\theta$ .)<sup>43</sup> As before, the period loss (A.21) becomes unboundedly large as  $|p_t^1|$  approaches the upper bound (3.2), so that we conclude once again that robustly optimal policy must satisfy this bound.

The robustly optimal linear policy is again a choice of  $\{p_t\}$  (subject to the constraints discussed above) to minimize the objective (3.8), except that the period loss is now given by (A.21). The solution to this problem can again be characterized as in section A.2. The first-order conditions for optimal CB policy are again of the form (A.14) – (A.15), but now the partial derivatives of  $L$  are equal to

$$\begin{aligned} L_\pi(\pi_t; p_t; s_t) &= \pi_t + \frac{\lambda}{\kappa^2} \frac{\pi_t - u_t - \kappa x^* - \beta p_t^0}{\Delta_t^2}, \\ L_0(\pi_t; p_t; s_t) &= -\beta \frac{\lambda}{\kappa^2} \frac{\pi_t - u_t - \kappa x^* - \beta p_t^0}{\Delta_t^2}, \\ L_1(\pi_t; p_t; s_t) &= 2 \frac{\beta^2}{\theta} \left( \frac{\lambda}{\kappa^2} \right)^2 \frac{(\pi_t - u_t - \kappa x^* - \beta p_t^0)^2}{\Delta_t^3} p_t^1. \end{aligned}$$

---

<sup>43</sup>The discussion here assumes that policy satisfies the bound (3.2). As above, if the inequality (3.2) is reversed, or it holds with equality and policy does not also satisfy (A.9), the malevolent agent chooses an extreme distortion of beliefs and the CB's period loss is unboundedly large. In the case that (3.2) holds with equality and (A.9) is precisely satisfied, the malevolent agent is indifferent among a set of distortions which are not all equally bad for the CB. In this case, the set of possible outcomes that might equally be chosen by the malevolent agent include ones in which the period loss of the CB is unboundedly large. Because we assume that the policy analyst should be concerned with the worst-case possible outcomes for her principal, it is reasonable to assign the CB loss function a value of  $+\infty$  in this case as well. The CB loss function is then defined for all possible linear policies, and it is evident that a robustly optimal policy will necessarily satisfy (3.2).



Condition (A.14) again implies that the process  $\{p_t^0\}$  must satisfy a stochastic linear difference equation of the form (A.16), except that we now define

$$A(L) \equiv \beta - \left(1 + \beta + \frac{\kappa^2 \bar{\Delta}^2}{\lambda}\right) L + L^2.$$

(The definition of  $v_t$  remains as before.) The unique stationary solution to (A.16) is again of the form (A.17), and again  $0 < \mu < 1$ , but now  $\mu$  is the smaller root of the characteristic equation

$$P(\mu) \equiv \beta\mu^2 - \left(1 + \beta + \frac{\kappa^2 \bar{\Delta}^2}{\lambda}\right) \mu + 1 = 0 \quad (\text{A.22})$$

rather than of equation (3.13) in the text. Again one finds that a stationary solution exists regardless of the value of  $\bar{p}^1$ , allowing one to define the ergodic measure  $\rho$  associated with any given value of  $\bar{p}^1$ ; and once again the stationary fluctuations occur around a long-run mean inflation rate of zero. Since (A.22) is the same as (3.13), except that  $\bar{\Delta}$  is replaced by  $\bar{\Delta}^2$ , one again observes that the root  $\mu$  is monotonically decreasing in  $\bar{\Delta}$ , so that a concern for robustness increases the persistence of the fluctuations in  $\{p_t^0\}$ . In the special case of i.i.d. cost-push shocks, (A.17) again reduces to (A.18), establishing conditions (3.11) given in the text.

Condition (A.15) again implies a condition of the form  $f(\bar{p}^1) = 0$ , but now we define

$$f(\bar{p}^1) \equiv 2\frac{\beta^2}{\theta} \left(\frac{\lambda}{\kappa^2}\right)^2 \frac{c}{\bar{\Delta}^3} \bar{p}^1 + \bar{p}^1 + \frac{\lambda}{\kappa^2 \bar{\Delta}^2} [\bar{p}^1 + b] = 0. \quad (\text{A.23})$$

Here the definition of  $c$  is as before, and one can again show that  $c > 0$  for all admissible values of  $\bar{p}^1$ . One can again show that in the case of any finite  $\theta$ ,  $f(0) < 0$ , while  $f(\hat{p}^1) > 0$  for any  $\hat{p}^1$  close enough to (while still less than) the bound  $\hat{p}^1$ . Hence a solution  $0 < \bar{p}^1 < \hat{p}^1$  exists, corresponding to a robustly optimal linear policy of the kind described in the text. Figure 4 shows how  $\bar{p}^1$  varies with parameters in the case of the same numerical values as in Figure 1. Note that the qualitative conclusions from this exercise are the same as before; quantitatively, we see that a given degree of concern for robustness (*i.e.*, a given value of  $\theta$ ) reduces the optimal sensitivity of inflation to the cost-push disturbances to an even greater extent when the CB is assumed to evaluate worst-case outcomes using (2.2).

One can also establish once again that the robustly optimal linear policy must satisfy (3.14). When evaluated at  $p^1 = \mu\sigma_u$ , the second two terms in the function

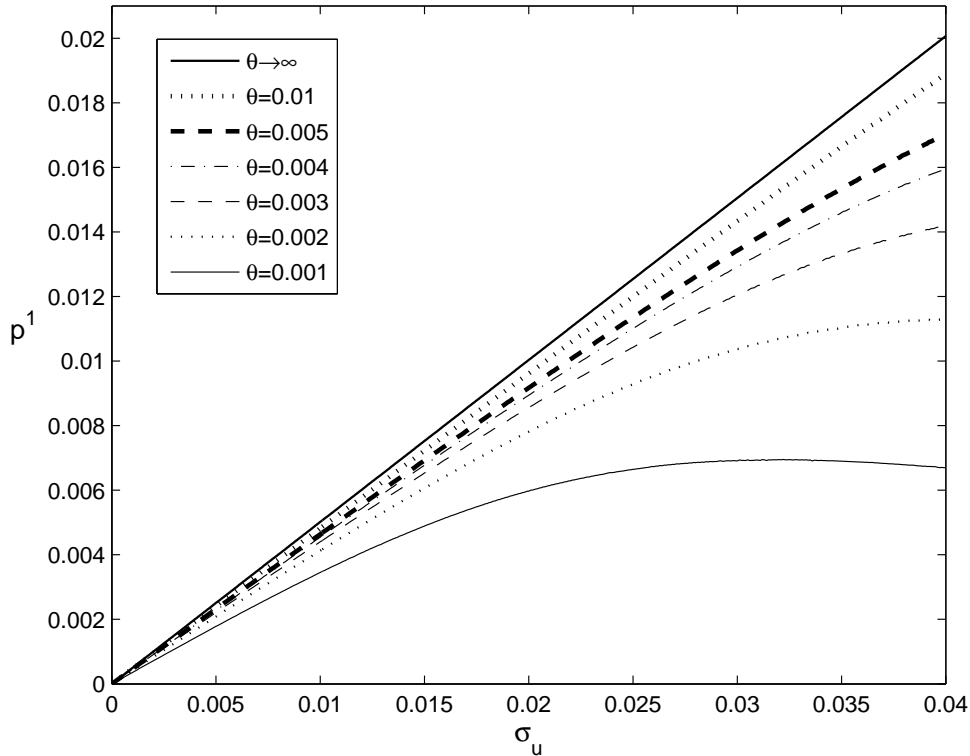


Figure A.1: Variation of  $\bar{p}^1$  with  $\sigma_u$ , under alternative degrees of concern for robustness, for the alternative CB objective.

defined in (A.23) are again equal to

$$-\frac{\lambda}{\kappa^2 \Delta} P(\mu) \sigma_u = 0,$$

where  $P(\mu)$  is now defined in (A.22), so that once again  $f(\mu\sigma_u) > 0$ . Thus it again follows that there exists a solution to (A.23) with  $0 < \bar{p}^1 < \mu\sigma_u$ . Since the cumulative log price increase due to a one-standard-deviation cost-push shock is again given by (3.15), this result again implies that with a concern for robustness, the initial price increase due to a positive cost-push shock is not merely eventually undone, but is eventually reversed in sign, under the optimal linear policy.

In the case of discretionary policy, use of the alternative CB objective similarly leaves the qualitative conclusions announced in the text unchanged. The same argument as in the text can be used to show that  $\bar{\pi}(s_t)$  is the solution to equation (4.2), except that now the function  $L$  is the one defined in (A.21), rather than the one

defined in (A.10). The solution is then given by

$$\bar{\pi}(s_t) = \frac{\lambda}{\kappa^2 \bar{\Delta} + \lambda} [\kappa x^* + u_t + \beta \bar{p}^0(s_t)].$$

rather than by (4.6). Again we have a linear solution of the form (4.3), and the expressions given in (4.7) – (4.8) continue to apply, if the factor  $\bar{\Delta}$  is replaced by  $\bar{\Delta}^2$  in each equation.

Because  $\bar{\Delta}^2$  is a decreasing, strictly concave function of  $\bar{p}^1$  just as  $\bar{\Delta}$  is, and similarly varies from 1 to 0 as  $\bar{p}^1$  increases from 0 to  $\hat{p}^1$ , the qualitative conclusions obtained in Appendix A.4 continue to apply in the case of the alternative objective; thus, for example, a large enough value of  $\kappa$  will once again result in the three regions shown in Figure 4. Details of the corresponding calculations for the alternative case are omitted.

#### A.4 Existence and Stability of Robust Linear MPE

A robust linear MPE corresponds to a pair  $(\bar{p}^1, \bar{\Delta})$  that satisfy equations (fixpt1) – (fixpt2), with  $\bar{\Delta} > 0$  so that (3.2) is satisfied. Equivalently, we are looking for solutions to the two equations in the interval  $0 < \bar{p}^1 < \hat{p}^1$ , where  $\hat{p}^1$  is defined by (A.20).

If we write these equations as  $\bar{\Delta} = \Delta_1(\bar{p}^1)$  and  $\bar{\Delta} = \Delta_2(\bar{p}^1)$  respectively, we observe that  $\Delta_1(p)$  is a decreasing, strictly concave function for all  $p > 0$ , while  $\Delta_2(p)$  is a decreasing, strictly convex function over the same domain. Moreover,  $\Delta_1(p) < \Delta_2(p)$  for all small enough  $p > 0$  (as  $\Delta_2(p) \rightarrow +\infty$  as  $p \rightarrow 0$ ), and also for all large enough  $p$  (as  $\Delta_1(p) \rightarrow -\infty$  as  $p \rightarrow +\infty$ ). Hence there are either *no* intersections of the two curves with  $\bar{p}^1 > 0$ , or *two* intersections, or a single intersection at a point of tangency between the two curves.

The slopes of the two curves are furthermore given by

$$\begin{aligned} \Delta'_1(p) &= -2 \frac{\beta^2 \lambda}{\theta \kappa^2} p, \\ \Delta'_2(p) &= -\frac{\lambda \sigma_u}{\kappa^2 p^2}. \end{aligned}$$

From these expressions one observes that  $\Delta'_2(p)$  is less than, equal to, or greater than  $\Delta'_1(p)$  according to whether  $p$  is less than, equal to, or greater than  $\tilde{p}^1$ , where

$$\tilde{p}^1 \equiv \left( \frac{\theta \sigma_u}{\beta^2 \kappa^2} \right)^{1/3} > 0.$$

From this it follows that there are two intersections if and only if  $\Delta_2(\bar{p}^1) < \Delta_1(\bar{p}^1)$ , which holds if and only if  $\sigma_u < \sigma_u^*$ , where  $\sigma_u^*$  is defined as in (4.11).<sup>44</sup> Similarly, the two curves are tangent to each other if and only if  $\sigma_u = \sigma_u^*$ ; in this case, the unique intersection is at  $\bar{p}^1 = \hat{p}^1$ . And finally, the two curves fail to intersect if and only if  $\sigma_u > \sigma_u^*$ .

It remains to consider how many of these intersections occur in the interval  $0 < \bar{p}^1 < \hat{p}^1$ . One notes that there is exactly one solution in that interval (and hence a unique robust linear MPE) if and only if  $\Delta_2(\hat{p}^1) < 0$ , which holds if and only if  $\sigma_u < \hat{p}^1$ . When  $\sigma_u = \hat{p}^1$  exactly,  $\Delta_2(\hat{p}^1) = \Delta_1(\hat{p}^1) = 0$ , and the curves intersect at  $\bar{p}^1 = \hat{p}^1$ . This is the larger of two solutions for  $\bar{p}^1$  if and only if

$$\Delta_1'(\hat{p}^1) < \Delta_2'(\hat{p}^1), \quad (\text{A.24})$$

which holds if and only if  $\lambda\kappa^2 < 2$ . In this case, as  $\sigma_u$  is increased further, the larger of the two solutions for  $\bar{p}^1$  decreases with  $\sigma_u$ , so that there are two solutions in the interval  $(0, \hat{p}^1)$ , until  $\sigma_u = \sigma_u^*$ , and the two solutions collapse into one, as the curves are tangent. (Note that  $\sigma_u^* > \hat{p}^1$ .) For still larger values of  $\sigma_u$ , there is no intersection, as explained in the previous paragraph.

If instead,  $\lambda/\kappa^2 = 2$  exactly, then the curves are tangent when  $\sigma_u = \hat{p}^1$  (which in this case is also equal to  $\sigma_u^*$ ). At this point the only intersection occurs at  $\bar{p}_1 = \hat{p}^1$  (which fails to satisfy condition (3.2)), and for larger values of  $\sigma_u$  there are no intersections. Finally, if  $\lambda/\kappa^2 > 2$ , then the inequality in (A.24) is reversed, and when  $\sigma_u = \hat{p}^1$ , the intersection at  $\bar{p}^1 = \hat{p}^1$  is the smaller of the two solutions. (The smaller solution approaches  $\hat{p}^1$  from below as  $\sigma_u$  increases to  $\hat{p}^1$ .) In this case, there are no solutions  $\bar{p}^1 < \hat{p}^1$  when  $\sigma_u = \hat{p}^1$ . As  $\sigma_u$  increases further, the smaller solution continues to increase with  $\sigma_u$ , so that even for values of  $\sigma_u$  that continue to be less than or equal to  $\sigma_u^*$  (so that the curves continue to intersect), there are no solutions with  $\bar{p}^1 < \hat{p}^1$ . And for still larger values of  $\sigma_u$ , there are again no solutions at all. Hence in each case, the number of solutions is as described in the text.

The “expectational stability” analysis proposed in the text involves the properties of the map

$$\Phi(p) \equiv \Delta_2^{-1}(\Delta_1(p)).$$

Formally, a fixed point  $\bar{p}^1$  of  $\Phi$  (which corresponds to an intersection of the two curves studied above) is expectationally stable if and only if there exists a neighborhood  $P$

---

<sup>44</sup>It is useful to note that this definition implies that  $\sigma_u^* \geq \hat{p}^1$ , with equality only if  $\kappa^2/\lambda = 1/2$ .

of  $\bar{p}^1$  such that

$$\lim_{n \rightarrow \infty} \Phi^n(p) = \bar{p}^1$$

for any  $p \in P$ . Our observations above about the functions  $\Delta_1(\cdot), \Delta_2(\cdot)$  imply that  $\Phi(\cdot)$  is a monotonically increasing function. Hence a fixed point  $\bar{p}^1$  is stable if and only if  $\Phi'(\bar{p}^1) < 1$ .

The above definition implies that

$$\Phi'(p) = \frac{\Delta_1'(p)}{\Delta_2'(\Delta_2^{-1}(\Delta_1(p)))} > 0.$$

Evaluated at a fixed point of  $\Phi$ , this reduces to

$$\Phi'(\bar{p}^1) = \frac{\Delta_1'(\bar{p}^1)}{\Delta_2'(\bar{p}^1)}.$$

Hence the stability condition is satisfied if and only if

$$\Delta_2'(\bar{p}^1) < \Delta_1'(\bar{p}^1) < 0. \quad (\text{A.25})$$

Because of the concavity of  $\Delta_1(\cdot)$  and the convexity of  $\Delta_2(\cdot)$ , this condition necessarily holds at the fixed point with the smaller value of  $\bar{p}^1$ , and not at the higher value. Hence in Figure 4, it is the upper (dashed) branch of solutions that is expectationally unstable, while the lower (solid) branch of solutions is stable. We therefore conclude that regardless of the other parameter values, there is exactly one expectationally stable robust linear MPE for all values of  $\sigma_u$  below some positive critical value, and no robust linear MPE for values of  $\sigma_u$  greater than or equal to that value.

Finally, let us consider the way in which  $\bar{p}^1$  changes as  $\theta$  is reduced (indicating that a broader range of NRE beliefs are considered possible). Letting  $\bar{p}^1$  be implicitly defined by the equation

$$\Delta_1(\bar{p}^1) = \Delta_2(\bar{p}^1),$$

the implicit function theorem implies that

$$\frac{d\bar{p}^1}{d\theta} = -\frac{\partial\Delta_1/\partial\theta}{\Delta_1' - \Delta_2'}. \quad (\text{A.26})$$

It follows from (A.25) that in the case of an expectationally stable MPE, the denominator of the fraction in (A.26) is positive. We also observe that

$$\frac{\partial\Delta_1}{\partial\theta} = \frac{\beta^2}{\theta^2} \frac{\lambda}{\kappa^2} (\bar{p}^1)^2 > 0,$$

so that the numerator is positive as well, and hence  $\bar{p}^1$  decreases as  $\theta$  increases. This means that  $\bar{p}^1$  increases as the CB's concern for robustness increases (corresponding to a lower value of  $\theta$ , up until the point where there ceases to any longer be a robust linear MPE at all. In that case, as discussed in the text, we can think of the equilibrium sensitivity of inflation to cost-push shocks as being *unbounded*; so the conclusion that greater concern for robustness leads to greater sensitivity of inflation to cost-push shocks extends, in a looser sense, to that case as well.

# References

- Clarida, Richard, Jordi Gali and Mark Gertler, “The Science of Monetary Policy: A New Keynesian Perspective,” *Journal of Economic Literature* 37: 1661-1707 (1999).
- Evans, George W., and Seppo Honkapohja, *Learning and Expectations in Macroeconomics*, Princeton: Princeton University Press, 2001.
- Gaspar, Vitor, Frank Smets and David Vestin, “Optimal Monetary Policy under Adaptive Learning,” presentation at the Annual Meeting, European Economics Association, Amsterdam, August 2005.
- Gilboa, Itzhak, and David Schmeidler, “Maxmin Expected Utility with Nonunique Prior,” *Journal of Mathematical Economics* 18: 141-153 (1989).
- Hansen, Lars Peter and Thomas J. Sargent, “Robust Estimation and Control under Commitment,” unpublished, University of Chicago, June 2005a.
- — and — —, “Recursive Robust Estimation and Control without Commitment,” unpublished, University of Chicago, July 2005b.
- — and — —, *Robustness*, unpublished, University of Chicago, September 2005c.
- —, — —, Gauhar Turmuhambetova, and Noah Williams, “Robust Control and Misspecification,” unpublished, University of Chicago, September 2005.
- Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini, “Ambiguity Aversion, Robustness, and the Variational Representation of Preferences,” unpublished, Università Bocconi, July 2004.
- —, — —, and — —, “Dynamic Variational Preferences,” unpublished, Università Bocconi, July 2005.
- Orphanides, Athanasios, and John C. Williams, “Imperfect Knowledge, Inflation Expectations, and Monetary Policy,” in B.S. Bernanke and M. Woodford, eds., *The Inflation Targeting Debate*, Chicago: University of Chicago Press, 2005a.
- — and — —, “Monetary Policy with Imperfect Knowledge,” unpublished, Federal Reserve Board, August 2005b.

Woodford, Michael, *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton: Princeton University Press, 2003.

— —, “An Example of Robustly Optimal Monetary Policy with Near-Rational Expectations,” presentation at the Annual Meeting, European Economics Association, Amsterdam, August 2005.