Journal of Monetary Economics 59 (2012) 468–487

Contents lists available at SciVerse ScienceDirect

Journal of Monetary Economics

journal homepage: www.elsevier.com/locate/jme

Robustly optimal monetary policy in a microfounded New Keynesian model ☆

Klaus Adam [a], Michael Woodford [b,*]

[a] *University of Mannheim, Germany*
[b] *Columbia University, United States*

ARTICLE INFO

Available online 5 June 2012

JEL classification:
D81
D84
E52

Keywords:
Robust control
Near-rational expectations
Belief distortions
Target criterion

ABSTRACT

We consider optimal monetary stabilization policy in a New Keynesian model with explicit microfoundations, 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 close enough to model-consistency. We show how to characterize robustly optimal policy without restricting consideration a priori to a particular parametric family of candidate policy rules. We show that robustly optimal policy can be implemented through commitment to a target criterion involving only the paths of inflation and a suitably defined output gap, but that a concern for robustness requires greater resistance to surprise increases in inflation than would be considered optimal if one could count on the private sector to have "rational expectations."

1. Introduction

A central issue in macroeconomic policy analysis is the need to take account of the likely changes in *people's expectations about the future* that should result from the adoption of one policy or another. The most common approach to confronting this issue in analyses of macroeconomic policy over the past 30–40 years has been to hypothesize "rational" (or model-consistent) expectations on the part of all economic agents. For any contemplated policy, this involves determining the outcome (the predicted state-contingent evolution of the economy) that would represent a rational expectations equilibrium (REE) according to one's model and the policy under consideration. One then compares the outcomes under the different REE associated with the different policies, in order to decide which policy is preferable.

While this is certainly a hypothesis of appealing simplicity and generality, it is also a very strong and restrictive hypothesis. There are important reasons to doubt the reliability of policy evaluation exercises that are based – or at least that are solely based – on models that assume that whatever policy may be adopted, everyone in the economy will necessarily and immediately understand the consequences of the policy commitment in exactly the same way as the policy analyst does.

Even if one is willing to suppose that people are thoroughly rational and possess extraordinary abilities at calculation, it is hardly obvious that they must forecast the economy's evolution in the same way as an economist's own model forecasts

☆ Prepared for the Carnegie–Rochester–NYU Conference on Public Policy, "Robust Macroeconomic Policy," November 11–12, 2011. We thank Pierpaolo Benigno, Roberto Colacito, Lars Hansen, Burton Hollifield, Luigi Paciello and Tack Yun for helpful comments, and the European Research Council (Starting Grant no. 284262) and the Institute for New Economic Thinking for research support.

* Corresponding author.
E-mail address: mw2230@columbia.edu (M. Woodford).

0304-3932/$ - see front matter
http://dx.doi.org/10.1016/j.jmoneco.2012.05.003

it; for even if the economist's model is completely correct, there will be many other possible models of the economy's probabilistic evolution that are (i) internally consistent, and (ii) not plainly contradicted by observations of the economy's evolution in the past. This is especially true given the relatively short sample of past observations that will be available in practice.

The assumption of model consistent expectations is an even more heroic one in the case that a change in policy is contemplated, relative to the pattern of conduct of policy with which people had experience in the past. Indeed, it is likely that there exist many internally consistent economic models that are consistent with the probabilistic evolution of the economy in the past, which make rather different predictions regarding the effects of a change in policy, while it is hard or even impossible to identify from past data which of these models is actually correct. Hence one should be cautious about drawing strong conclusions about the character of desirable policies solely on the basis of an analysis that maintains the assumption of model consistent expectations.

Here we explore a different approach, under which the policy analyst should not pretend to be able to model the precise way in which people will form expectations, if a particular policy is adopted. Instead, under our recommended approach, the policy analyst recognizes that the public's beliefs might be anything in a certain set of possible beliefs, satisfying the requirements of (i) internal consistency, and (ii) not being *too grossly inconsistent* with what actually happens in equilibrium, when people act on the basis of those beliefs. These requirements reduce to the familiar assumption of model-consistent ("rational") expectations if the words "not too grossly inconsistent" are replaced by "completely consistent."[1] The weakening of the standard requirement of model-consistent expectations is motivated by the recognition that it makes sense to expect people's beliefs to take account of patterns in their environment that are clear enough to be obvious after even a modest period of observation, while there is much less reason to expect them to have rejected an alternative hypothesis that is not easily distinguishable from the true model after only a series of observations of modest length.

Under this approach, the economic analyst's model will associate with each contemplated policy not a unique prediction about what people in the economy will expect under that policy, but rather a *range of possible forecasts*; and there will correspondingly be a range of possible predictions for economic outcomes under the policy, rather than a unique prediction. In essence, it is proposed that one's economic model be used to place *bounds* on what can occur under a given policy, rather than expecting a point prediction. In the spirit of the literatures on "ambiguity aversion" and on "robust control",[2] and following Woodford (2010), we then suppose that the economic analyst chooses a policy that ensures as high as possible a value of one's objective under *any* of the set of possible outcomes associated with that policy.[3] Under a particular precise definition of what it means for expectations to be sufficiently close to model-consistency, this criterion again allows a unique policy to be recommended. It will, however, differ in general from the one that would be selected if one were confident that people's expectations would have to be fully consistent with the predictions of one's model.

The present paper illustrates the consequences of this alternative approach to policy selection in the context of a New Keynesian model of the tradeoff between inflation and output stabilization that is based on explicit choice-theoretic foundations. As in Woodford (2010), we adopt a particular interpretation of the requirement of "near-rational expectations" by supposing that the policy analyst assumes that people's beliefs will be absolutely continuous with respect to the measure implied by her own model[4] and that their beliefs will not be too different from the prediction of her model, where the distance is measured by a relative entropy criterion. A policy can then be said to be "robustly optimal" if it guarantees as high as possible a value of the policymaker's objective, under any of the subjective beliefs consistent with the above criterion.

This non-parametric way of specifying the range of beliefs that are "close enough" to the policy analyst's own beliefs is based on the approach to bounding possible model mis-specifications in the robust policy analysis of Hansen and Sargent (2005).[5] It has the advantage, in our view, of allowing us to be fairly agnostic about the nature of the possible alternative beliefs that may be entertained by the public. In addition, it retains a high degree of theoretical parsimony: it simply defines a one-parameter family of robustly optimal policies, indexed by a parameter that can be taken to measure the policy analyst's degree of concern for robustness to possible departures from model-consistent expectations.

[1] The more general proposal is termed an assumption of "near-rational expectations" in Woodford (2010).

[2] See Hansen and Sargent (2008, 2011) for a discussion of these ideas and their application to decision problems arising in macroeconomics.

[3] Alternatively, the policy analyst ensures that a certain lower bound for the policy objective is achieved under as broad as possible a range of possible departures from model-consistent expectations.

[4] This implies that people correctly identify zero-probability events as having zero probability, though they may differ in the probability they assign to events that occur with positive probability according to her model.

[5] Our use of this measure of departure from model-consistent expectations is somewhat different from theirs, however. Hansen and Sargent assume a policy analyst who is herself uncertain that her model is precisely correct as a description of the economy; when the expectations of other economic agents are an issue in the analysis, these are typically assumed to share the policy analyst's model, and her concerns about mis-specification and preference for robustness as well. We are instead concerned about potential discrepancies between the views of the policy analyst and those of the public; and the potential departures from model-consistent beliefs on the part of the public are not assumed to reflect a concern for robustness on their part. In Benigno and Paciello (2010), instead, optimal policy is computed under the assumption that members of the public are concerned about the robustness of *their own* decisions, and the policymaker correctly understands the way that this distorts their actions (relative to what the policymaker believes would be optimal for them). Hansen and Sargent (unpublished) consider a similar exercise.

Using a model with explicit choice-theoretic foundations, the present paper re-examines the policy conclusions reached in Woodford (2010) who incorporates robustness concerns by modifying the linear quadratic approximation to the optimal policy problem that emerges from an analysis of the New Keynesian model under rational expectations.[6] Such a re-examination is of interest because it is not obvious that the proposed modification of these equations can similarly be justified as a local approximation when expectations are allowed to deviate from model-consistent ones.[7] Here we derive exact, nonlinear equations that characterize a robustly optimal policy commitment in the context of our microfounded model, before log-linearizing those equations to provide a local linear approximation to the solution to those equations; this is intended to guarantee that the linear approximations that are eventually relied upon to obtain our final, practical characterizations are invoked in an internally consistent manner.

The present analysis also generalizes the approach in Woodford (2010) who optimizes policy over only a family of linear policy rules of a particular restrictive form. While this restriction is known not to matter in the case of an analysis of optimal policy in the log-linear approximate model under rational expectations,[8] it is not obvious that there may not be advantages to alternative types of rules when one allows for departures for rational expectations. For this reason, we consider here robustly optimal policy choice from among a much more flexibly specified class of policies, including policies that allow for an explicit response to measures or indicators of private-sector expectations. While this has no advantage under an RE analysis, since no such discrepancy can ever exist in an REE, one might expect it to be desirable for policy to respond to observed departures of public expectations from those that the central bank regards as correct. Yet, to the extent that our criterion for robustness is simply one of ensuring that the highest possible lower bound for welfare across alternative ''near-rational'' beliefs is achieved, we find that there is no benefit from expanding the set of candidate policy commitments to include ones that are explicitly dependent on private-sector expectations.[9] But it is an important advance of the current analysis that this can be shown rather than simply being assumed.

Our findings confirm a key conclusion of Woodford (2010): a concern for robustness requires greater resistance to surprise increases in inflation than would be considered optimal if one could count on the private sector to have rational expectations.[10] While we discuss a variety of policy commitments that can each equally be characterized as ''robustly optimal'' – as they ensure the same lower bound for expected utility across possible ''near-rational'' belief distortions – all of them involve dynamics of inflation in response to shocks, under the worst-case private-sector beliefs, with smaller surprise movements in inflation.[11] A particularly convenient representation of a policy commitment is in terms of a *target criterion*—a linear relationship between the paths of inflation and the output gap that the central bank is committed to maintain regardless of the evolution of shocks. We show that the robustly optimal target criterion requires larger output-gap deviations to justify allowing a given size of inflation surprise than the criterion (derived in Benigno and Woodford, 2005) that would be optimal under rational expectations.[12]

In Section 2, we explain our general approach to the characterization of robustly optimal policy. In addition to introducing our proposed definition of ''near-rational expectations,'' this section explains in general terms how it is possible for us to characterize robustly optimal policy without having to restrict the analysis to a parametric family of candidate policy rules, as is done in Woodford (2010). Section 3 then sets out the structure of the microfounded New Keynesian model, showing how the model's exact structural relations are modified by the allowance for distorted private-sector expectations. Section 4 begins the analysis of robustly optimal policy in the New Keynesian model by characterizing an evolution of the economy that represents an upper bound on what can possibly be achieved. Section 5 provides an approximate analysis of the upper-bound dynamics by log-linearizing the exact conditions established in Section 4; Section 6 then shows that (at least up to the linear approximation introduced in Section 5) the upper-bound dynamics are attainable by a variety of policies, and hence solve the robust policy problem stated earlier. Section 7 then briefly considers a stronger form of robustness, and Section 8 concludes.

2. Robustly optimal policy: preliminaries

Here we first describe the general strategy of the approach that we use to characterize robustly optimal policy. These general ideas are then applied to a specific New Keynesian model in Section 3.

[6] Benigno and Woodford (2005) shows that for the case of rational expectations, the linear-quadratic policy problem assumed in Woodford (2010) provides a local approximation to the dynamics under an optimal policy commitment in a microfounded New Keynesian model.

[7] Benigno and Paciello (2010) criticize the analysis of Woodford (2010) on this ground. Tack Yun has raised the same issue, in a discussion of Woodford (2010) at a conference at the Bank of Korea.

[8] See, e.g., Clarida et al. (1999), or Section 1 in Woodford (2011).

[9] In Section 7, we discuss a stronger form of robustness that is more difficult to achieve, and argue that robustness in this stronger sense *would* require a commitment to respond to fairly direct measures of belief distortions.

[10] See Section 5.4 for a comparison of the present results with the earlier findings in Woodford (2010).

[11] See Section 5.3 for the characterization of inflation dynamics under the worst-case beliefs.

[12] See Section 6, especially Eq. (71).

2.1. The robustly optimal policy problem

Our general strategy for characterizing robustly optimal policy can be usefully explained in a fairly abstract setting, before turning to an application of the approach in the context of a specific model. In particular, we wish to explain how it is possible to characterize robustly optimal policy without restricting consideration to a particular parametric family of policy rules, as is done in Woodford (2010).

Let us suppose in general terms that a policymaker cares about economic outcomes that can be represented by some vector x of endogenous variables, the values of which will depend both on policy and on private-sector belief distortions, with the latter parameterized by some vector m.[13] Among the determinants of x are a vector of structural equations, that we write as

$$F(x,m)=0. \tag{1}$$

We assume that Eq. (1) are insufficient to completely determine the vector x, under given belief distortions m, so that the policymaker has a non-trivial choice.

We further assume that in the absence of any concern for possible belief distortions on the part of the private sector, i.e., if it were possible to be confident that private-sector beliefs would coincide with his own, the policymaker would wish to achieve as high a value as possible of some objective $W(x)$. In the application below, this objective will correspond to the expected utility of the representative household. In the presence of a concern for robustness, we instead assume, following Hansen and Sargent (2005) and Woodford (2010), that alternative policies are evaluated according to the value of

$$\min_{m\in M}[W(x)+\theta V(m)], \tag{2}$$

where the minimization is over the set of all possible belief distortions M; $V(m)\geq 0$ is a measure of the size of the belief distortions, equal to zero only in the case of beliefs that agree precisely with those of the policymaker; $\theta>0$ is a coefficient that indexes the policymaker's degree of concern about potential belief distortions; and (2) is evaluated taking into account the way in which belief distortions affect the determination of x. Here a small value of θ implies a great degree of concern for robustness, while a large value of θ implies that only modest departures from model-consistent expectations are considered plausible. In the limit as $\theta\rightarrow\infty$, criterion (2) reduces to $W(x)$, and the rational expectations analysis is recovered.[14]

More specifically, let us suppose that the policymaker must choose a policy commitment c from some set C of feasible policy commitments. Our goal is to show that we can obtain results about robustly optimal policy that do not depend on the precise specification of the set C; for now, we assume that there exists such a set, but we make no specific assumption about what its boundaries may be. We only make two general assumptions about the nature of the set C. First, we assume that each of the commitments in the set C can be defined independently of what the belief distortions may be.[15] And second, we shall require that for any $c\in C$, there exists an equilibrium outcome for any choice of $m\in M$.

We thus rule out policy commitments that would imply non-existence of equilibrium for some $m\in M$, and thereby situations in which one might be tempted to conclude that belief distortions must be of a particular type under a given policy commitment, simply because no other beliefs would be consistent with existence of equilibrium. Instead of assuming that private-sector beliefs will necessarily be consistent with some equilibrium that allows the intended policy to be carried out, we assume that it is the responsibility of the policymaker to choose a policy commitment that can be executed (so that an equilibrium exists in which it is fulfilled), regardless of the beliefs that turn out to be held by the private sector. Thus, if under certain beliefs, the policy would have to be modified on ground of infeasibility, then a credible description of the policy commitment should specify that the outcome will be different in the case of those beliefs.[16]

Note that the set C may involve many different types of policy commitments. For example, it may include policy commitments that depend on the history of exogenous shocks; commitments that depend on the history of endogenous variables, as is the case with Taylor rules; and commitments regarding relationships between endogenous variables, as is the case with so-called targeting rules. Also, the endogenous variables in terms of which the policy commitment is expressed may include asset prices (futures prices, forward prices, etc.) that are often treated by central banks as indicators of private-sector expectations, as long as the requirement is satisfied that the policy commitment must be consistent with belief distortions of an arbitrary form.

In order to define the robustly optimal decision problem of the policymaker, we further specify an *outcome function* that identifies the equilibrium outcome x associated with a given policy commitment and a given belief distortion m.

[13] In Section 2.3 we discuss a particular approach to the parameterization of belief distortions, but our general remarks here do not rely on it.

[14] Adam (2004) shows that the modified objective function (2) assumed for the case with a concern for robustness can be interpreted as inducing infinite risk aversion over a subset of the possible belief distortions. Again, the size of this subset depends inversely on the robustness parameter θ.

[15] As is made more specific in the application below, we specify policy commitments by equations involving the endogenous and exogenous variables x, and not explicitly involving the belief distortions m. But of course the endogenous variables referred to in the rule will typically also be linked by structural equations that involve the belief distortions.

[16] Alternatively, instead of ruling out commitments that give rise to non-existence of equilibrium under some belief distortions, it is equivalent to allow for such commitments and to assign a value of $-\infty$ to the policymaker's objective when an equilibrium does not exist.

Definition 1. The economic outcomes associated with belief distortions m and commitments c are given by an outcome function

$$O : M \times C \to X,$$

with the property that for all $m \in M$ and $c \in C$, the outcome $O(m,c)$ and m jointly constitute an equilibrium of the model. In particular, the outcome function must satisfy

$$F(O(m,c),m) = 0,$$

for all $m \in M$ and $c \in C$.

Here we have not been specific about what we mean by an "equilibrium," apart from the fact that (1) must be satisfied. In the context of the specific model presented in the next section, equilibrium has a precise meaning. For purposes of the present discussion, it does not actually matter how we define equilibrium; only the definition of the outcome function matters for our subsequent discussion.[17]

Note also that we do not assume that there is necessarily a *unique* equilibrium associated with each policy commitment c and belief distortion m. We simply suppose that the policymaker's robust policy problem can be defined relative to some assumption about which equilibrium should be selected in order to evaluate a given policy. For example, consistent with the desire for robustness, one might specify that the outcome function $O(c,m)$ selects the *worst* of the equilibria, in the sense of yielding the lowest value for $W(x)$ consistent with the pair (c,m). Our approach to the characterization of robustly optimal policy, however, does not depend on such a specification; it can also be used to determine the robustly optimal policy for a policymaker who is willing to assume that the *best* equilibrium will occur, among those consistent with the given belief distortion.

We are now in a position to define *the robustly optimal policy problem* as the choice of a policy commitment to solve

$$\max_{c \in C} \min_{m \in M} \Lambda(m,c), \tag{3}$$

where

$$\Lambda(m,c) \equiv W(O(m,c)) + \theta V(m).$$

2.2. An upper bound on what policy can robustly achieve

We shall now determine an upper bound for the economic outcomes that robustly optimal policy can achieve in the decision problem (3), that does not depend on the choice of the set C of feasible commitments or the outcome function $O(\cdot,\cdot)$. We proceed in three incremental steps.

First, we use the min–max inequality (see Appendix A.1 for a proof) to obtain

$$\max_{c \in C} \min_{m \in M} \Lambda(m,c) \le \min_{m \in M} \max_{c \in C} \Lambda(m,c). \tag{4}$$

This inequality captures the intuitively obvious fact that it is no disadvantage to be the second mover in the "game".

Second, using the right-hand side in (4), we free the policymaker from the restriction to choose commitments from the strategy space C and from the restrictions imposed by the outcome function $O(\cdot,\cdot)$. Instead, we allow the policymaker to choose directly the preferred economic outcomes x consistent with an equilibrium. This yields

$$\min_{m \in M} \max_{c \in C} \Lambda(m,c) \le \min_{m \in M} \max_{x \in X} [W(x) + \theta V(m)] \quad \text{s.t.} : F(x,m) = 0, \tag{5}$$

where the constraint $F(x,m)=0$ captures the restrictions required for x to be an equilibrium.[18]

In a third step, we define a *Lagrangian optimization problem* associated with problem (5)

$$\min_{m \in M} \max_{x \in X} L(m,x,\gamma), \tag{6}$$

where L is the Lagrange function

$$L(m,x,\gamma) \equiv W(x) + \theta V(m) + \gamma F(x,m),$$

and γ is a vector of Lagrange multipliers. We will now state conditions under which the outcome of the Lagrangian problem (6) generates weakly higher utility to the policymaker than problem (5). Under these conditions it will also be the case that the solution of the Lagrangian problem represents an upper bound on what policy can achieve in the robustly optimal policy problem (3).

Suppose we have found a point (m^*,x^*,γ^*) and the Lagrange function has a saddle at this point, i.e., satisfies

$$L(m^*,x,\gamma^*) < L(m^*,x^*,\gamma^*) \quad \forall x \ne x^*, \tag{7a}$$

[17] If the set of equations (1) is not a complete set of requirements for x to be an equilibrium, this only has the consequence that the upper-bound outcome defined below might not be a tight enough upper bound; it does not affect the validity of the assertion that it provides an upper bound.

[18] The constraint represents a restriction on the choice of the second mover, i.e., the policymaker choosing x.

$$L(m,x^*,\gamma^*) > L(m^*,x^*,\gamma^*) \quad \forall m \neq m^*, \tag{7b}$$

$$L(m^*,x^*,\gamma) \geq L(m^*,x^*,\gamma^*) \quad \forall \gamma. \tag{7c}$$

Appendix A.1 then proves the following result:

Proposition 1. *Suppose* (m^*,x^*,γ^*) *satisfies the saddle point conditions* (7) *and let* (x^R,m^R) *denote the solution of the robustly optimal policy problem* (3). *Then* (x^*,m^*) *is an equilibrium and*

$$W(x^R)+\theta V(m^R) \leq W(x^*)+\theta V(m^*).$$

The solution to the Lagrangian optimization problem thus delivers an upper bound on what policy can achieve in the robustly optimal policy problem, provided the saddle-point conditions hold.

Assuming differentiability, it follows from conditions (7a) and (7b) that the solution to the Lagrangian problem necessarily satisfies the first-order conditions

$$W_x(x^*)+\gamma^* F_x(x^*,m^*) = 0, \tag{8}$$

$$\theta V_m(m^*)+\gamma^* F_m(x^*,m^*) = 0. \tag{9}$$

Moreover, condition (7c) holds if and only if

$$F(x^*,m^*) = 0. \tag{10}$$

Conditions (8)–(10) represent necessary conditions that allow us to generate candidate solutions for the Lagrangian optimization problem. If a candidate solution satisfies (7a) and (7b), then Proposition 1 implies that one has found an upper bound to the value of the robustly optimal policy problem (3).[19] For simplicity we refer to the solution of the Lagrangian problem as the "upper-bound solution" in the remainder of the paper.

2.3. Distorted private sector expectations

We next discuss our approach to the parameterization of belief distortions, and the cost function $V(m)$. At this point it becomes necessary to specify that our analysis concerns dynamic models in which information is progressively revealed over time, at a countably infinite sequence of successive decision points.

Let $(\Omega,\mathcal{B},\mathcal{P})$ denote a standard probability space with Ω denoting the set of possible realizations of an exogenous stochastic disturbance process $\{\xi_0,\xi_1,\xi_2,\ldots\}$, $\mathcal{B}$ the σ-algebra of Borel subsets of Ω, and $\mathcal{P}$ a probability measure assigning probabilities to any set $B \in \mathcal{B}$. We consider a situation in which the policy analyst assigns probabilities to events using the probability measure $\mathcal{P}$ but fears that the private sector may make decisions on the basis of a potentially different probability measure denoted by $\widehat{\mathcal{P}}$.

We let E denote the policy analyst's expectations induced by $\mathcal{P}$ and $\widehat{E}$ the corresponding private sector expectations associated with $\widehat{\mathcal{P}}$. A first restriction on the class of possible distorted measures that the policy analyst is assumed to consider – part of what we mean by the restriction to "near-rational expectations" – is the assumption that the distorted measure $\widehat{\mathcal{P}}$, when restricted to events over any finite horizon, is absolutely continuous with respect to the correspondingly restricted version of the policy analyst's measure $\mathcal{P}$.

The Radon–Nikodym theorem then allows us to express the distorted private sector expectations of some $t+j$ measurable random variable X_{t+j} as

$$\widehat{E}[X_{t+j}|\xi^t] = E\left[\frac{\mathcal{M}_{t+j}}{\mathcal{M}_t}X_{t+j}|\xi^t\right],$$

for all $j \geq 0$ where ξ^t denotes the partial history of exogenous disturbances up to period t. The random variable $\mathcal{M}_{t+j}$ is the Radon–Nikodym derivative, and completely summarizes belief distortions.[20] The variable $\mathcal{M}_{t+j}$ is measurable with respect to the history of shocks ξ^{t+j}, non-negative and is a martingale, i.e., satisfies

$$E[\mathcal{M}_{t+j}|\omega^t] = \mathcal{M}_t,$$

for all $j \geq 0$. Defining

$$m_{t+1} = \frac{\mathcal{M}_{t+1}}{\mathcal{M}_t},$$

one step ahead expectations based on the measure $\widehat{\mathcal{P}}$ can be expressed as

$$\widehat{E}[X_{t+1}|\xi^t] = E[m_{t+1}X_{t+1}|\xi^t],$$

[19] Condition (7c) is implied by the necessary condition (10).
[20] See Hansen and Sargent (2005) for further discussion.

where m_{t+1} satisfies

$$E[m_{t+1}|\xi^t]=1 \quad \text{and} \quad m_{t+1}\geq 0. \tag{11}$$

This representation of the distorted beliefs of the private sector is useful in defining a measure of the distance of the private-sector beliefs from those of the policy analyst. As discussed in Hansen and Sargent (2005), the relative entropy

$$R_t = E_t[m_{t+1}\log m_{t+1}]$$

is a measure of the distance of (one-period-ahead) private-sector beliefs from the policymaker's beliefs with a number of appealing properties.

We wish to extend this measure of the size of belief distortions to an infinite-horizon economy with a stationary structure. In the kind of model with which we are concerned, the policy objective in the absence of a concern for robustness is of the form

$$W(x)\equiv E_0\left[\sum_{t=0}^{\infty}\beta^t U(x_t)\right], \tag{12}$$

for some discount factor $0<\beta<1$, where $U(\cdot)$ is a time-invariant function, and x_t is a vector describing the real allocation of resources in period t. Correspondingly, we propose to measure the overall degree of distortion of private-sector beliefs by a discounted criterion of the form

$$V(m)\equiv E_0\left[\sum_{t=0}^{\infty}\beta^{t+1}m_{t+1}\log m_{t+1}\right], \tag{13}$$

as in Woodford (2010). This is a discounted sum of the one-period-ahead distortion measures $\{R_t\}$. We assign relative weights to the one-period-ahead measures R_t for different dates and different states of the world in this criterion that match those of the other part of the policy objective (12). Use of this cost function implies that the policymaker's degree of concern for robustness (relative to other stabilization objectives) remains constant over time, regardless of past history.

Hansen and Sargent (2005) appear to use a different cost function, but this is because they consider a problem in which a decisionmaker is concerned about the possible inaccuracy of *her own* probability beliefs. In their problem, the decisionmaker's basic objective is of the form

$$W^{HS}(x)\equiv \widehat{E}_0\left[\sum_{t=0}^{\infty}\beta^t U(x_t)\right], \tag{14}$$

instead of (12), as she wishes to maximize expected utility under the *correct* probabilities, which may be different from those implied by her baseline model. They correspondingly define a discounted measure of belief distortions

$$V^{HS}(m)\equiv \widehat{E}_0\left[\sum_{t=0}^{\infty}\beta^{t+1}m_{t+1}\log m_{t+1}\right] = E_0\left[\sum_{t=0}^{\infty}\beta^{t+1}M_{t+1}\log m_{t+1}\right] \tag{15}$$

instead of (13). As in their analysis, our worst-case belief distortions minimize a discounted sum of terms of the form $U(x_t)+\beta\theta R_t$, with a relative weight θ that is time-invariant.[21] This allows us to obtain a characterization of robustly optimal policy with a stationary form, which simplifies the presentation of our results below.

It may be asked why we do not assume an objective of the form (14) in our case, in which case it would also be appropriate to assume a cost function (15) for belief distortions. This would imply a desire to maximize the expected utility of the representative household *as evaluated by that household* when forecasting the consequences of its actions, whether the policymaker agrees with those beliefs or not. We instead assume a paternalistic objective: the policymaker wishes to maximize people's *true welfare*, whether they understand it correctly or not.

There are arguments to be made for either objective in a normative analysis. The non-paternalistic case, however, can be dealt with easily. In that case, the policymaker's problem is equivalent to the choice of a policy under the assumption that a rational-expectations equilibrium must result, but with uncertainty about the true probabilities of the stochastic disturbances (assumed to be correctly understood by the private sector). Since the rational-expectations analysis of Giannoni and Woodford (2010) has already shown that there exists a form of policy commitment (commitment to an optimal target criterion) that achieves a welfare-optimal equilibrium *regardless* of the stochastic process assumed for the exogenous disturbances, this would also be a robustly optimal policy commitment under the non-paternalistic objective. In the remainder of this paper, we instead deal with the more complex case in which the policymaker's objective is paternalistic.[22]

We now apply these results to a specific New Keynesian DSGE model of the options for monetary stabilization policy.

[21] The point of the discount factor in (15) is clearly to make this relative weight time-invariant.

[22] See Hansen and Sargent (2011) for additional discussion of alternative possible robustly optimal policy problems in the context of a dynamic New Keynesian model.

3. A New Keynesian model with distorted private sector expectations

We shall begin by deriving the exact structural relations of a New Keynesian model that is completely standard, except that the private sector holds potentially distorted expectations. The exposition here follows and extends (Woodford, 2011), who writes the exact structural relations in a recursive form for the case with model-consistent expectations.

3.1. Model structure

The economy is made up of identical infinite-lived households, each of which seeks to maximize

$$U \equiv \widehat{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\tilde{u}(C_t;\xi_t) - \int_0^1 \tilde{v}(H_t(j);\xi_t)\, dj \right], \tag{16}$$

subject to a sequence of flow budget constraints[23]

$$P_t C_t + B_t \le \int_0^1 w_t(j) P_t H_t(j)\, dj + B_{t-1}(1+i_{t-1}) + \Sigma_t + T_t,$$

where $\widehat{E}_0$ is the common distorted expectations held by consumers conditional on the state of the world in period t_0, C_t an aggregate consumption good which can be bought at nominal price P_t, $H_t(j)$ is the quantity supplied of labor of type j and $\omega_t(j)$ the associated real wage, B_t nominal bond holdings, i_t the nominal interest rate, and ξ_t is a vector of exogenous disturbances, which may include random shifts of either of the functions $\tilde{u}$ or $\tilde{v}$. The variable T_t denotes lump sum taxes levied by the government and Σ_t denotes profits accruing to households from the ownership of firms.

The aggregate consumption good is a Dixit–Stiglitz aggregate of consumption of each of a continuum of differentiated goods

$$C_t \equiv \left[\int_0^1 c_t(i)^{(\eta-1)/\eta}\, di \right]^{\eta/(\eta-1)}, \tag{17}$$

with an elasticity of substitution equal to $\eta > 1$. We further assume isoelastic functional forms

$$\tilde{u}(C_t;\xi_t) \equiv \frac{C_t^{1-\tilde{\sigma}^{-1}} \overline{C}_t^{\tilde{\sigma}^{-1}}}{1-\tilde{\sigma}^{-1}}, \tag{18}$$

$$\tilde{v}(H_t;\xi_t) \equiv \frac{\lambda}{1+v} H_t^{1+v} \overline{H}_t^{-v}, \tag{19}$$

where $\tilde{\sigma}$, $v > 0$, and $\{\overline{C}_t, \overline{H}_t\}$ are bounded exogenous disturbance processes which are both among the exogenous disturbances included in the vector ξ_t.

Each differentiated good is supplied by a single monopolistically competitive producer; there is a common technology for the production of all goods, in which (industry-specific) labor is the only variable input

$$y_t(i) = A_t f(h_t(i)) = A_t h_t(i)^{1/\phi}, \tag{20}$$

where A_t is an exogenously varying technology factor, and $\phi > 1$. The Dixit–Stiglitz preferences (17) imply that the quantity demanded of each individual good i will equal[24]

$$y_t(i) = Y_t \left(\frac{p_t(i)}{P_t} \right)^{-\eta}, \tag{21}$$

where Y_t is the total demand for the composite good defined in (17), $p_t(i)$ is the (money) price of the individual good, and P_t is the price index,

$$P_t \equiv \left[\int_0^1 p_t(i)^{1-\eta}\, di \right]^{1/(1-\eta)}, \tag{22}$$

corresponding to the minimum cost for which a unit of the composite good can be purchased in period t. Total demand is given by

$$Y_t = C_t + g_t Y_t, \tag{23}$$

where g_t is the share of the total amount of composite good purchased by the government, treated here as an exogenous disturbance process.

[23] We abstract from state-contingent assets in the household budget constraint because the representative agent assumption implies that in equilibrium there will be no trade in these assets.

[24] In addition to assuming that household utility depends only on the quantity obtained of C_t, we assume that the government also cares only about the quantity obtained of the composite good defined by (17), and that it seeks to obtain this good through a minimum-cost combination of purchases of individual goods.

3.2. Household optimality conditions

Each household maximizes utility by choosing state contingent sequences $\{C_t, H_t(j), B_t\}$ taking as given the processes for $\{P_t, w_t(j), i_t, \Sigma_t, T_t\}$ and the preference shocks. The first order conditions give rise to an optimal labor supply relation

$$w_t(j) = \frac{\tilde{v}_h(H_t(j);\xi_t)}{\tilde{u}_c(C_t;\xi_t)}, \tag{24}$$

and a consumption Euler equation

$$\tilde{u}_C(C_t;\xi_t) = \beta \widehat{E}_t\left[\tilde{u}_C(C_t;\xi_t)\frac{1+i_t}{\Pi_{t+1}}\right], \tag{25}$$

which characterize optimal household behavior.

3.3. Optimal price setting by firms

The producers in each industry fix the prices of their goods in monetary units for a random interval of time, as in the model of staggered pricing introduced by Calvo (1983) and Yun (1996). Let $0 \le \alpha < 1$ be the fraction of prices that remain unchanged in any period. A supplier that changes its price in period t chooses its new price $p_t(i)$ to maximize

$$\widehat{E}_t \sum_{T=t}^{\infty} \alpha^{T-t} Q_{t,T} \Pi(p_t(i), p_T^j, P_T; Y_T, \xi_T), \tag{26}$$

where $\widehat{E}_t$ is the distorted expectations of price setters conditional on time t information, which are assumed identical to the expectations held by consumers, $Q_{t,T}$ is the stochastic discount factor by which financial markets discount random nominal income in period T to determine the nominal value of a claim to such income in period t, and α^{T-t} is the probability that a price chosen in period t will not have been revised by period T. In equilibrium, this discount factor is given by

$$Q_{t,T} = \beta^{T-t}\frac{\tilde{u}_c(C_T;\xi_T)}{\tilde{u}_c(C_t;\xi_t)}\frac{P_t}{P_T}. \tag{27}$$

Profits are equal to after-tax sales revenues net of the wage bill. Sales revenues are determined by the demand function (21), so that (nominal) after-tax revenue equals

$$(1-\tau_t)p_t(i)Y_t\left(\frac{p_t(i)}{P_t}\right)^{-\eta}.$$

Here τ_t is a proportional tax on sales revenues in period t; $\{\tau_t\}$ is treated as an exogenous disturbance process, taken as given by the monetary policymaker. We assume that τ_t fluctuates over a small interval around a non-zero steady-state level $\overline{\tau}$. We allow for exogenous variations in the tax rate in order to include the possibility of "pure cost-push shocks" that affect equilibrium pricing behavior while implying no change in the efficient allocation of resources.

The real wage demanded for labor of type j is given by Eq. (24) and firms are assumed to be wage-takers. Substituting the assumed functional forms for preferences and technology, the function

$$\Pi(p,p^j,P;Y,\xi) \equiv (1-\tau)pY(p/P)^{-\eta} - \lambda P\left(\frac{p}{P}\right)^{-\eta\phi}\left(\frac{p^j}{P}\right)^{-\eta\phi v}\overline{H}^{-v}\left(\frac{Y}{A}\right)^{1+\omega}\left(\frac{(1-g)Y}{\overline{C}}\right)^{1/\tilde{\sigma}}, \tag{28}$$

then describes the after-tax nominal profits of a supplier with price p, in an industry with common price p^j, when the aggregate price index is equal to P and aggregate demand is equal to Y. Here $\omega \equiv \phi(1+v)-1 > 0$ is the elasticity of real marginal cost in an industry with respect to industry output. The vector of exogenous disturbances ξ_t now includes A_t, g_t and τ_t, in addition to the preference shocks $\overline{C}_t$ and $\overline{H}_t$.

Each of the suppliers that revise their prices in period t chooses the same new price p_t^*, that maximizes (26). Note that supplier i's profits are a concave function of the quantity sold $y_t(i)$, since revenues are proportional to $y_t(i)^{(\eta-1)/\eta}$ and hence concave in $y_t(i)$, while costs are convex in $y_t(i)$. Moreover, since $y_t(i)$ is proportional to $p_t(i)^{-\eta}$, the profit function is also concave in $p_t(i)^{-\eta}$. The first-order condition for the optimal choice of the price $p_t(i)$ is the same as the one with respect to $p_t(i)^{-\eta}$; hence the first-order condition with respect to $p_t(i)$

$$\widehat{E}_t \sum_{T=t}^{\infty} \alpha^{T-t} Q_{t,T} \Pi_1(p_t(i), p_T^j, P_T; Y_T, \xi_T) = 0,$$

is both necessary and sufficient for an optimum. The equilibrium choice p_t^* (which is the same for each firm in industry j) is the solution to the equation obtained by substituting $p_t(i) = p_t^j = p_t^*$ into the above first-order condition.

Under the assumed isoelastic functional forms, the optimal choice has a closed-form solution

$$\frac{p_t^*}{P_t} = \left(\frac{K_t}{F_t}\right)^{1/(1+\omega\eta)}, \tag{29}$$

where F_t and K_t are functions of current aggregate output Y_t, the current exogenous state ξ_t, and the expected future evolution of inflation, output, and disturbances, defined by

$$F_t \equiv \widehat{E}_t \sum_{T=t}^{\infty} (\alpha\beta)^{T-t} f(Y_T;\xi_T)\left(\frac{P_T}{P_t}\right)^{\eta-1}, \tag{30}$$

$$K_t \equiv \widehat{E}_t \sum_{T=t}^{\infty} (\alpha\beta)^{T-t} k(Y_T;\xi_T)\left(\frac{P_T}{P_t}\right)^{\eta(1+\omega)}, \tag{31}$$

where

$$f(Y;\xi) \equiv (1-\tau)\overline{C}^{\tilde{\sigma}^{-1}}(Y(1-g))^{-\tilde{\sigma}^{-1}}Y, \tag{32}$$

$$k(Y;\xi) \equiv \frac{\eta}{\eta-1}\lambda\phi\frac{1}{A^{1+\omega}\overline{H}^{\nu}}Y^{1+\omega}. \tag{33}$$

Relations (30) and (31) can instead be written in the recursive form

$$F_t = f(Y_t;\xi_t) + \alpha\beta\widehat{E}_t[\Pi_{t+1}^{\eta-1}F_{t+1}], \tag{34}$$

$$K_t = k(Y_t;\xi_t) + \alpha\beta\widehat{E}_t[\Pi_{t+1}^{\eta(1+\omega)}K_{t+1}], \tag{35}$$

where $\Pi_t \equiv P_t/P_{t-1}$.[25]

The price index then evolves according to a law of motion

$$P_t = [(1-\alpha)p_t^{*1-\eta} + \alpha P_{t-1}^{1-\eta}]^{1/(1-\eta)}, \tag{36}$$

as a consequence of (22). Substitution of (29) into (36) implies that equilibrium inflation in any period is given by

$$\frac{1-\alpha\Pi_t^{\eta-1}}{1-\alpha} = \left(\frac{F_t}{K_t}\right)^{(\eta-1)/(1+\omega\eta)}. \tag{37}$$

Eqs. (34), (35) and (37) jointly define a short-run aggregate supply relation between inflation and output, given the current disturbances ξ_t, and expectations regarding future inflation, output, and disturbances.

3.4. Summary of the model equations and equilibrium definition

For the subsequent analysis it will be helpful to express the model in terms of the endogenous variables $(K_t, F_t, Y_t, i_t, \Delta_t, m_t)$ only, where m_t is the belief distortions of the private sector and

$$\Delta_t \equiv \int_0^1 \left(\frac{p_t(i)}{P_t}\right)^{-\eta(1+\omega)} di \geq 1, \tag{38}$$

a measure of price dispersion at time t. The vector of exogenous disturbances is given by $\xi_t = (A_t, g_t, \tau_t, \overline{C}_t, \overline{H}_t)'$.

We begin by expressing expected household utility (evaluated under the objective measure $\mathcal{P}$) in terms of these variables. Inverting the production function (20) to write the demand for each type of labor as a function of the quantities produced of the various differentiated goods, and using the identity (23) to substitute for C_t, where g_t is treated as exogenous, it is possible to write the utility of the representative household as a function of the expected production plan $\{y_t(i)\}$. One thereby obtains

$$U \equiv E_0 \sum_{t=0}^{\infty} \beta^t \left[u(Y_t;\xi_t) - \int_0^1 v(y_t^j;\xi_t)dj\right], \tag{39}$$

where

$$u(Y_t;\xi_t) \equiv \tilde{u}(Y_t(1-g_t);\xi_t)$$

and

$$v(y_t^j;\xi_t) \equiv \tilde{v}(f^{-1}(y_t^j/A_t);\xi_t).$$

In this last expression we make use of the fact that the quantity produced of each good in industry j will be the same, and hence can be denoted y_t^j; and that the quantity of labor hired by each of these firms will also be the same, so that the total demand for labor of type j is proportional to the demand of any one of these firms.

[25] It is evident that (30) implies (34); but one can also show that processes that satisfy (34) each period, together with certain bounds, must satisfy (30). Since we are interested below only in the characterization of bounded equilibria, we can omit the statement of the bounds that are implied by the existence of well-behaved expressions on the right-hand sides of (30) and (31), and treat (34) and (35) as necessary and sufficient for processes $\{F_t, K_t\}$ to measure the relevant marginal conditions for optimal price-setting.

One can furthermore express the relative quantities demanded of the differentiated goods each period as a function of their relative prices, using (21). This allows us to write the utility flow to the representative household in the form

$$U(Y_t,\Delta_t;\xi_t) \equiv u(Y_t;\xi_t) - v(Y_t;\xi_t)\Delta_t.$$

Hence we can express the household objective (39) as

$$U = E_0 \sum_{t=0}^{\infty} \beta^t U(Y_t,\Delta_t;\xi_t). \tag{40}$$

Here $U(Y,\Delta;\xi)$ is a strictly concave function of Y for given Δ and ξ, and a monotonically decreasing function of Δ given Y and ξ.

Using this notation, the consumption Euler equation (25) can be expressed as

$$u_Y(Y_t;\xi_t) = \beta E_t\left[m_{t+1} u_Y(Y_{t+1};\xi_{t+1}) \frac{1+i_t}{\Pi_{t+1}} \frac{1-g_t}{1-g_{t+1}}\right]. \tag{41}$$

Using (37) to substitute for the variable Π_t Eqs. (34) and (35) can be expressed as

$$F_t = f(Y_t;\xi_t) + \alpha\beta E_t[m_{t+1}\phi_F(K_{t+1},F_{t+1})], \tag{42}$$

$$K_t = k(Y_t;\xi_t) + \alpha\beta E_t[m_{t+1}\phi_K(K_{t+1},F_{t+1})], \tag{43}$$

where the functions ϕ_F, ϕ_K are both homogeneous degree 1 functions of K and F.

Because the relative prices of the industries that do not change their prices in period t remain the same, one can use (36) to derive a law of motion for the price dispersion term Δ_t of the form

$$\Delta_t = h(\Delta_{t-1},\Pi_t),$$

where

$$h(\Delta,\Pi) \equiv \alpha\Delta\Pi^{\eta(1+\omega)} + (1-\alpha)\left(\frac{1-\alpha\Pi^{\eta-1}}{1-\alpha}\right)^{(\eta(1+\omega))/(\eta-1)}.$$

This is the source of welfare losses from inflation or deflation. Using once more (37) to substitute for the variable Π_t one obtains

$$\Delta_t = \tilde{h}(\Delta_{t-1},K_t/F_t). \tag{44}$$

Eqs. (41)–(44) represent four constraints on the equilibrium paths of the six endogenous variables $(Y_t,F_t,K_t,\Delta_t,i_t,m_t)$. For a given sequence of belief distortions m_t satisfying restriction (11) there is thus one degree of freedom left, which can be determined by monetary policy. We are now in a position to define the equilibrium with distorted private sector expectations:

Definition 2 (*DEE*). A distorted expectations equilibrium (DEE) is a stochastic process for $\{Y_t,F_t,K_t,\Delta_t,i_t,m_t\}_{t=0}^{\infty}$ satisfying Eqs. (11) and (41)–(44).

4. Upper bound in the New Keynesian model

We shall now formulate the Lagrangian optimization problem (6) for the nonlinear New Keynesian model with distorted private sector expectations, and derive the nonlinear form of the necessary conditions (8)–(10).

The Lagrangian game (6) for the New Keynesian model is given by

$$\min_{\{m_{t+1}\}_{t=0}^{\infty}} \max_{\{Y_t,F_t,K_t,\Delta_t\}_{t=0}^{\infty}} E_0 \sum_{t=0}^{\infty} \beta^t \begin{bmatrix} U(Y_t,\Delta_t;\xi_t) + \theta\beta m_{t+1}\log m_{t+1} \\ +\gamma_t(\tilde{h}(\Delta_{t-1},K_t/F_t)-\Delta_t) \\ +\Gamma_t'[z(Y_t;\xi_t)+\alpha\beta m_{t+1}\Phi(Z_{t+1})-Z_t] \\ +\beta\psi_t(m_{t+1}-1) \end{bmatrix} + \alpha\Gamma_{-1}'\Phi(Z_0), \tag{45}$$

where γ_t, Γ_t, ψ_t denote Lagrange multipliers and we used the shorthand notation

$$Z_t \equiv \begin{bmatrix} F_t \\ K_t \end{bmatrix}, \quad z(Y;\xi) \equiv \begin{bmatrix} f(Y;\xi) \\ k(Y;\xi) \end{bmatrix}, \quad \Phi(Z) \equiv \begin{bmatrix} \phi_F(K,F) \\ \phi_K(K,F) \end{bmatrix}, \tag{46}$$

and added the initial pre-commitment $\alpha\Gamma_{-1}'\Phi(Z_0)$ to obtain a time-invariant solution. The Lagrange multiplier vector Γ_t is associated with constraints (42) and (43) and given by $\Gamma_t' = (\Gamma_{1t},\Gamma_{2,t})$. The multiplier γ_t relates to Eq. (44) and the multiplier ψ_t to constraint (11). We also eliminated the interest rate and the constraint (41) from the problem. Under the assumption that the zero lower bound on nominal interest rates is not binding, constraint (41) imposes no restrictions on the path of the other variables.[26]

The path for the nominal interest rates can thus be computed ex-post using the solution for the remaining variables and Eq. (41).

The nonlinear FOCs for the policymaker (8) are then given by

$$U_Y(Y_t,\Delta_t;\xi_t)+\Gamma_t' z_Y(Y_t;\xi_t)=0, \tag{47}$$

$$-\gamma_t \tilde{h}_2(\Delta_{t-1},K_t/F_t)\frac{K_t}{F_t^2}-\Gamma_{1t}+\alpha m_t \Gamma_{t-1}' D_1(K_t/F_t)=0, \tag{48}$$

$$\gamma_t \tilde{h}_2(\Delta_{t-1},K_t/F_t)\frac{1}{F_t}-\Gamma_{2t}+\alpha m_t \Gamma_{t-1}' D_2(K_t/F_t)=0, \tag{49}$$

$$U_\Delta(Y_t,\Delta_t;\xi_t)-\gamma_t+\beta E_t[\gamma_{t+1}\tilde{h}_1(\Delta_t,K_{t+1}/F_{t+1})]=0, \tag{50}$$

for all $t\geq 0$. The nonlinear FOC (9) defining the worst-case belief distortions takes the form

$$\theta(\log m_t+1)+\alpha\Gamma_{t-1}'\Phi(Z_t)+\psi_{t-1}=0, \tag{51}$$

for all $t\geq 1$. Above, $\tilde{h}_i(\Delta,K/F)$ denotes the partial derivative of $\tilde{h}(\Delta,K/F)$ with respect to its ith argument, and $D_i(K/F)$ is the ith column of the matrix

$$D(Z)\equiv\begin{bmatrix}\partial_F\phi_F(Z) & \partial_K\phi_F(Z)\\ \partial_F\phi_K(Z) & \partial_K\phi_K(Z)\end{bmatrix}. \tag{52}$$

Since the elements of $\Phi(Z)$ are homogeneous degree 1 functions of Z, the elements of $D(Z)$ are all homogenous degree 0 functions of Z, and hence functions of K/F only. Thus we can alternatively write $D(K/F)$. Finally, the structural Eq. (10) are given by Eqs. (42)–(44). This completes the description of the necessary conditions Eqs. (8)–(10) for the New Keynesian model.

5. Locally optimal dynamics under the upper bound policy

We shall be concerned solely with optimal outcomes that involve small fluctuations around a deterministic optimal steady state. An *optimal steady state* is a set of constant values $(\overline{Y},\overline{Z},\overline{\Delta},\overline{\gamma},\overline{\Gamma},\overline{\psi},\overline{m})$ that solve the structural equations (42)–(44) and the FOCs (47)–(51) in the case that $\xi_t=\overline{\xi}$ at all times and initial conditions consistent with the steady state are assumed. We now compute the steady-state, then derive the local dynamics implied by these FOCs and show that the saddle point conditions (7) are locally satisfied.

5.1. Optimal steady state

In a deterministic steady state, restriction (11) implies $\overline{m}=1$, so that the optimal steady state is the same as derived in Benigno and Woodford (2005) for the case with non-distorted private sector expectations. Specifically, it satisfies $\overline{F}=\overline{K}=(1-\alpha\beta)^{-1}k(\overline{Y};\overline{\xi})$, which implies $\overline{\Pi}=1$ (no inflation) and $\overline{\Delta}=1$ (zero price dispersion), and the value of $\overline{Y}$ is implicitly defined by

$$f(\overline{Y},\overline{\xi})=k(\overline{Y},\overline{\xi}).$$

Because $\tilde{h}_2(1,1)=0$ (the effects of a small non-zero inflation rate on the measure of price dispersion are of second order), conditions (48) and (49) reduce in the steady state to the eigenvector condition

$$\overline{\Gamma}'=\alpha\overline{\Gamma}'D(1). \tag{53}$$

Moreover, since when evaluated at a point where $F=K$

$$\frac{\partial\log(\phi_K/\phi_F)}{\partial\log K}=-\frac{\partial\log(\phi_K/\phi_F)}{\partial\log F}=\frac{1}{\alpha},$$

and we observe that $D(1)$ has a left eigenvector $[1\ -1]$, with eigenvalue $1/\alpha$; hence (53) is satisfied if and only if $\overline{\Gamma}_2=-\overline{\Gamma}_1$. Condition (47) provides then one additional condition to determine the magnitude of the elements of $\overline{\Gamma}_1$. It implies

$$U_Y(\overline{Y},1;\overline{\xi})+\overline{\Gamma}_1(f_Y(\overline{Y};\overline{\xi})-k_Y(\overline{Y};\overline{\xi}))=0. \tag{54}$$

[26] This assertion also depends on our assumption here that the central bank chooses its interest-rate operating target i_t with full information about the state of the economy at date t. Our results are extended to the (more realistic) case in which the central bank must act without complete knowledge of the current state in the working paper version of this paper.

Since $k_y - f_y = \omega + \tilde{\sigma}^{-1} > 0$ we have that

$$\overline{\Gamma}_1 > 0,$$

whenever $U_Y > 0$, i.e., whenever steady state output $\overline{Y}$ falls short of the first best or efficient steady state level $\overline{Y}^e$ defined as

$$U_Y(\overline{Y}^e, 1; \overline{\xi}) = 0.$$

In the limiting case $\overline{Y} \to \overline{Y}^e$ we have $\overline{\Gamma}_1 = 0$. Finally, condition (50) provides a restriction allowing to determine the steady state value of $\overline{\gamma}$

$$U_\Delta(\overline{Y}, 1; \overline{\xi}) - \overline{\gamma} + \beta \overline{\gamma} \tilde{h}_1(1,1) = 0.$$

Since $U_\Delta < 0$ and $\tilde{h}_1(1,1) = \alpha$, we have

$$\overline{\gamma} = \frac{U_\Delta(\overline{Y}, 1; \overline{\xi})}{(1 - \beta\alpha)} < 0.$$

5.2. Optimal dynamics

Let us define the endogenous variables

$$\pi_t \equiv \log \Pi_t,$$

$$\hat{m}_t \equiv \log m_t,$$

$$x_t \equiv \hat{Y}_t - \hat{Y}_t^*, \tag{55}$$

where x_t denotes the 'output gap' with $\hat{Y}_t = \log Y_t / \overline{Y}$, $\hat{Y}_t^* = \log Y_t^* / \overline{Y}$ and Y_t^* being the 'target level of output', which is a function of the exogenous disturbances only and implicitly defined as

$$U_Y(Y_t^*, 1; \xi_t) + \overline{\Gamma}' z_Y(Y_t^*; \xi_t) = 0. \tag{56}$$

The following proposition characterizes the log-linear local approximation to the dynamics implied by the nonlinear structural equations (42)–(44) and the nonlinear first-order conditions (47)–(51):

Proposition 2. *If initial price dispersion Δ_{-1} is small (of order $\mathcal{O}(\|\xi\|^2)$) and the initial precommitments such that $\Gamma_{1,0} = -\Gamma_{2,0} > 0$, then Eqs. (42)–(44) and (47)–(51) imply up to first order that*

$$\pi_t = \kappa x_t + \beta E_t \pi_{t+1} + u_t, \tag{57}$$

$$0 = \xi_\pi \pi_t + \lambda_x (x_t - x_{t-1}) + \xi_m \hat{m}_t, \tag{58}$$

$$\hat{m}_t = \lambda_m (\pi_t - E_{t-1}[\pi_t]). \tag{59}$$

The constants ($\kappa > 0, \xi_\pi, \xi_m, \lambda_x, \lambda_m$) are functions of the deep model parameters (explicit expressions are provided in Appendix A.2). In the empirically relevant case in which steady state output falls short of its efficient level ($\overline{Y} < \overline{Y}^e$) we have $\xi_\pi > 0, \xi_m > 0, \lambda_m > 0$; and if the steady-state output distortion is sufficiently small, $\lambda_x > 0$ as well.

The proof of the proposition is given in Appendix A.2. The disturbance u_t above denotes a 'cost-push' term and is defined as

$$u_t \equiv \kappa[\hat{Y}_t^* + u_\xi' \tilde{\xi}_t], \tag{60}$$

where u_ξ is defined in Eq. (80) in Appendix A.2. It is straightforward to generalize the above proposition to the case with larger degrees of initial price dispersion (Δ_{-1} of order $\mathcal{O}(\|\xi\|)$). As becomes clear from Appendix A.2, this would add additional deterministic dynamics to the optimal path. Also, in the case that the initial precommitments fail to imply the condition stated in the proposition, the results of the proposition would still become valid asymptotically, as the effects of the initial conditions vanishes with time.

The following proposition shows that the economic outcomes characterized by Proposition 2 indeed constitute a local solution to the upper-bound problem (5).

Proposition 3. *If steady state output falls short of its efficient level ($\overline{Y} < \overline{Y}^e$) and the steady state output distortions are sufficiently small, then the Lagrangian (45) locally satisfies the saddle point properties (7a)–(7b) at the solution implied by Eqs. (57)–(59).*

The proof of the proposition can be found in Appendix A.2.

5.3. The optimal inflation response to cost-push disturbances

In this section we derive a closed form solution for the optimal inflation response to a cost push disturbance, as implied by Eqs. (57)–(59). For simplicity, we assume that the evolution of the cost-push disturbances is described by

$$u_t = \rho u_{t-1} + \omega_t, \tag{61}$$

where $\rho \in [0,1)$ captures the persistence of the disturbance and ω_t is an *iid* innovation. We then use the relationship (59) to substitute for $\hat{m}_t$ in (58), and Eq. (57) to substitute for x_t. This delivers a second-order expectational difference equation describing the worst-case inflation evolution under a robustly optimal policy commitment

$$0 = \xi_\pi \pi_t + \frac{\lambda_x}{\kappa}(\pi_t - \beta E_t \pi_{t+1} - u_t - \pi_{t-1} + \beta E_{t-1}\pi_t + u_{t-1}) + \xi_m \lambda_m (\pi_t - E_{t-1}\pi_t).$$

We now consider the impulse response dynamics to an unexpected cost push shock ω_{t_0} in some period t_0 that are implied by this equation. Because of the linearity of our system, we can calculate the dynamic response to an individual shock independently of any assumptions about the shocks that occur in other periods, so let us consider the case in which no shocks have occurred in the past and none will occur in any later periods either; in this case we need only solve for the perfect-foresight dynamics after the occurrence of the one-time shock. We suppose, then, that we start from the deterministic steady state, so that the initial conditions are given by $\pi_{t_0-1} = E_{t_0-1}\pi_{t_0} = u_{t_0-1} = 0$. The previous equation then implies

$$0 = \left(\xi_\pi + \xi_m \lambda_m + \frac{\lambda_x}{\kappa}\right)\pi_{t_0} - \frac{\lambda_x}{\kappa}(\beta \pi_{t_0+1} + u_{t_0}), \tag{62}$$

$$0 = \left(\xi_\pi + \frac{\lambda_x(1+\beta)}{\kappa}\right)\pi_t - \frac{\lambda_x}{\kappa}(\beta \pi_{t+1} + \pi_{t-1} + u_t - u_{t-1}) \quad \text{for } t > t_0, \tag{63}$$

where the second equation applies for all $t > t_0$. (All variables in these equations refer to the expected values of the variables after the shock is realized in period t_0.)

The eigenvalues of the characteristic equation imply that Eq. (63) has a unique non-explosive solution for π_t $(t > t_0)$ for a given initial value π_{t_0} and a given bounded exogenous sequence for u_t. In the case that (as implied by (61)) $u_{t+j} = \rho^j u_t$ for all $j \geq 0$, so that at each date u_t is a sufficient statistic for the entire anticipated future evolution of the disturbance term, this solution takes the simple form

$$\pi_t = a\pi_{t-1} + bu_{t-1}, \tag{64}$$

where $0 < a < 1$ is the smaller of the two real roots of

$$\beta\mu^2 - (1 + \beta + \xi_\pi \kappa/\lambda_x)\mu + 1 = 0$$

and

$$b = -(1-\rho)a < 0.$$

Note that the coefficients a and b are independent of the policymaker's concern for robustness θ. Thus the optimal dynamics for $t > t_0$ depend in the same way on the lagged inflation rate and the path of the exogenous disturbance as in a pure RE analysis of the model. The result is different, though, for the initial period t_0 when inflation jumps unexpectedly in response to the shock.

Combining Eq. (61) with Eq. (64) for $t = t_0 + 1$ delivers a solution of the form $\pi_{t_0} = b_0 u_{t_0}$ for the initial impact of the shock, where

$$b_0 \equiv \frac{b + \beta^{-1}}{\frac{\kappa}{\lambda_x \beta}\left(\xi_\pi + \xi_m \lambda_m + \frac{\lambda_x}{\kappa}\right) - a}.$$

Note that the numerator and denominator of this fraction are both positive for all $\xi_m \lambda_m \geq 0$, so that $b_0 > 0$. With robustness concerns we have $\xi_m \lambda_m > 0$, so that the optimal immediate impact effect of the shock on inflation is smaller than under the RE analysis.

Thus we obtain an important result already mentioned in the Introduction: a concern for robustness to possible departures from model-consistent expectations makes it optimal for the inflation surprise resulting from a cost-push shock to be *smaller* than it would be under an optimal commitment assuming rational expectations. In the limiting case where robustness concerns increase without bound ($\theta \to 0$), we have $\xi_m \lambda_m \to \infty$, so that it becomes optimal to prevent any unexpected jump in inflation at all in response to a shock. (That is, under an optimal policy, inflation will be completely forecastable one period in advance.)

It follows that the cumulative price level response to a shock is given by

$$\sum_{t=t_0}^{\infty} \pi_t = \frac{b_0 u_{t_0}}{1-a} + \sum_{t=t_0+1}^{\infty} \frac{bu_t}{1-a} = \left[b_0 + \left(\frac{\rho}{1-\rho}\right)b\right]\frac{u_{t_0}}{1-a}.$$

In the absence of robustness concerns, this implies that $\sum_{t=0}^{\infty}\pi_t = 0$, so that cost-push shocks have no effect on the long-run price level under an optimal commitment. (This results in the familiar conclusion from the RE literature that price-level targeting is optimal.) Since a and b are independent of robustness concerns, but the initial response b_0 is dampened under robustness concerns, the term in square brackets is negative when robustness is taken into account. Hence robustness concerns make it optimal to plan to decrease the price level in the long run following a positive cost-push shock, and similarly to increase it following a negative shock.

Because of certainty-equivalence, the above results translate directly to the case with a random shock each period, as specified in (61). Under the upper-bound dynamics, in any period t_0, the conditional expectation $E_{t_0}\pi_t$ (for any $t \geq t_0$) depends linearly on u_{t_0} through precisely the coefficient obtained in the perfect-foresight calculation, so that the sequence of coefficients describes the impulse response function of inflation to a cost-push shock. The law of motion for inflation in the general case is given by

$$\pi_t = E_{t-1}\pi_t + (\pi_t - E_{t-1}\pi_t) = (a\pi_{t-1} + bu_{t-1}) + b_0(u_t - \rho u_{t-1}) = a\pi_{t-1} + b_0u_t + b_1u_{t-1}, \tag{65}$$

where $b_1 \equiv b - \rho b_0 < 0$. Thus inflation evolves according to the stationary ARMA(2,1) process

$$(1-aL)(1-\rho L)\pi_t = b_0\omega_t + b_1\omega_{t-1}.$$

5.4. Comparison with results in Woodford (2009)

As noted in the introduction, Woodford (2010) considers a similar problem, but assuming a quadratic loss function

$$\min E_0 \sum_{t=t_0}^{\infty} \beta^t[\pi_t^2 + \lambda(x_t - x^*)^2], \tag{66}$$

with coefficients λ, $x^* > 0$ for the policy objective, and a New Keynesian Phillips curve that depends on subjective private-sector expectations

$$\pi_t = \kappa x_t + \hat{E}_t\pi_{t+1} + u_t. \tag{67}$$

The structural relation (67) is assumed to be linear in the (potentially) distorted expectations, but when written in terms of the policymaker's expectation operator

$$\pi_t = \kappa x_t + E_t[m_{t+1}\pi_{t+1}] + u_t, \tag{68}$$

the structural relation includes a quadratic term.

It is known from the results in Benigno and Woodford (2005) that the characterization of the optimal policy commitment obtained from such a linear-quadratic analysis coincides with the linear approximation to the dynamics under an optimal policy commitment that can be derived (as in the present paper) by log-linearizing the exact equations that characterize an optimal commitment in a microfounded New Keynesian model.[27] Here we comment on the extent to which a similar justification for the linear-quadratic analysis is valid when policy is required to be robust to departures from model-consistent expectations.

In Woodford (2010), worst-case dynamics under the robustly optimal policy commitment are described by linear equations, as they are here, but the linearity is obtained not from a local linear approximation to the exact optimal dynamics, but rather as a consequence of only optimizing over a class of linear policy rules. The analysis in Woodford (2010) therefore leaves open the question of the extent to which nonlinear policy rules could improve upon the constrained-optimal policy characterized in that paper, while our present analysis leaves open the question of the extent to which the optimal policy commitment should be different in the case of larger shocks than those assumed in our local analysis. Hence we should not expect the results of the two analyses to coincide, except in the case to which both are intended to give a solution, which is the case of small enough shocks for terms other than those of first order in the amplitude of the shocks to be neglected.[28] Woodford (2010) also presents an explicit solution for the dynamics under robustly optimal policy only in the case of i.i.d. cost-push disturbances, corresponding to the special case $\rho = 0$ of the process (61) considered in the previous section.

We can, however, compare the results obtained here to those obtained in Woodford (2010) for the case $\rho = 0$ in the small-shock limit (i.e., the limiting values of the coefficients that describe the robustly optimal dynamics as $\sigma_u \to 0$). In that limiting case, the results presented in (2010) coincide with those derived here, with a suitable interpretation of the coefficients λ, x^* of the policy objective (66) in terms of the parameters of our microfounded model.

In Woodford (2010), as here, the dynamics of inflation under the robustly optimal policy commitment[29] are given by a law of motion of the form (65); in the earlier paper, the coefficient a is referred to as μ, the coefficient b_0 is referred to as $\overline{p}_1/\sigma_u$, and the coefficient b_1 (which is equal to $-a$ in the case that $\rho = 0$) is written as $-\mu$. The characteristic equation

[27] See Woodford (2011, Section 2) for further discussion of the relation between the two approaches.

[28] In fact, the results obviously do not coincide more generally, since the coefficients of the robustly optimal linear dynamics derived in Woodford (2010) are functions of the parameter σ_u, indicating the standard deviation of the "cost-push shocks," whereas they are independent of all shock variances in the local linear approximation calculated in this paper.

[29] Under the kind of policy assumed in Woodford (2010), the dynamics of inflation are determined solely by the policy commitment and are independent of private-sector belief distortions. As discussed in the next section, this is also one possible way of implementing the upper-bound dynamics in our model as well, though not the only one.

defining a in the present solution is furthermore seen to coincide with the quadratic equation defining μ in Woodford (2010) if the coefficient λ in that paper is defined as

$$\lambda \equiv \frac{\lambda_x \kappa}{\xi_\pi},$$

in terms of our current notation.[30] Moreover, the nonlinear equation that implicitly defines $\overline{p}_1$ in Woodford (2010) implies that $\overline{p}_1 \to 0$ as $\sigma_u \to 0$, but that the ratio $\overline{p}_1/\sigma_u$ converges to a non-zero limit. That limiting value is given by an equation identical to the one given above for b_0, if x^* is the positive quantity[31] such that

$$\left(\frac{\beta\lambda}{\kappa} x^*\right)^2 = \frac{\xi_m}{\xi_\pi} \lambda_m \theta > 0.$$

Hence with these identifications of the parameter values, the linear dynamics for inflation derived in Woodford (2010) are identical to those obtained here as a linear approximation to the upper-bound dynamics.

Hence the problem considered in Woodford (2010) has the same solution as the robustly optimal dynamics of our microfounded model, up to a linear approximation of these respective characterizations in the limiting case of small-enough exogenous disturbances. We have no reason, however, to expect that the characterization in Woodford (2010) of the way in which robustly optimal policy changes as σ_u is increased should also be correct for the microfounded model. There is no reason to expect even that the calculations in the earlier paper describe robustly optimal policy within the class of linear policy rules; for in this sort of calculation for the large-shock case, nonlinearities of the various structural equations become relevant, and we have no reason to suppose that the particular nonlinearity that is considered in Woodford (2010) – the effect of the distorted expectations in (68) – is the only that is quantitatively significant. But we leave the quantitative investigation of this issue for future work.

6. Implementing the upper bound

We now study whether a monetary policymaker can achieve the upper-bound solution characterized in the previous section, so that it represents the solution to the robustly optimal monetary policy problem (3). Since we have only characterized the upper-bound dynamics to a linear approximation, we similarly only show that certain policies result in dynamics that coincide with the upper-bound dynamics in this local linear approximation. We do show that local implementation is feasible, and present a variety of policy commitments, each of which would suffice for this purpose.

The result that we rely upon applies to policy commitments of the following form.

Assumption 1. Under policy commitment c, the policymaker commits to ensure that some relationship $c_t(\cdot) = 0$ holds every period, where for each t, $c_t(\cdot)$ is a function of the paths of the variables $\{\Pi_\tau, Y_\tau, i_\tau, \xi_\tau\}$ for $\tau \le t$, and there exists some neighborhood of the steady-state values of these variables such that the functions $c_t(\cdot)$ are all defined and twice continuously differentiable for all paths that remain forever within that neighborhood.

We shall furthermore seek a robustly optimal member of a class of rules of the following form.

Definition 3. In the case of any neighborhood $\mathcal{N}$ of the steady-state values of the endogenous variables (Π_t, Y_t, i_t); any bound $\|\xi\| > 0$ on the amplitude of the exogenous disturbances; and any class $\mathcal{M}$ of belief distortions, including all processes $\{m_{t+1}\}$ in which m_{t+1} remains forever within a certain neighborhood of 1; we define the class $\mathcal{C}$ of policy commitments as the set of all commitments c such that:

1. Assumption 1 is satisfied, in the case of exogenous disturbances satisfying the bound $\|\xi\|$ and paths of the endogenous variables remaining forever within neighborhood $\mathcal{N}$; and
2. for any belief distortions in the class $\mathcal{M}$, and any disturbance process satisfying the bound $\|\xi\|$, there exists at least one DEE in which the endogenous variables remain forever in the neighborhood $\mathcal{N}$.

Suppose furthermore that there exists a policy commitment with the following additional properties.

Assumption 2. The policy commitment c is consistent with the steady state in the case that all disturbance processes are at all times equal to their steady-state values. Moreover, a log-linear approximation of the sequence of policy commitments around the steady state is such that:

1. the linearized policy commitments are consistent with the log-linear approximation to the upper-bound solution (defined by Eqs. (57)–(59)), and
2. the linearized policy commitments imply a locally determinate equilibrium under rational expectations (i.e., there exist a bound on the amplitude of the exogenous disturbances and a neighborhood of the steady-state values of the

[30] Note that as long as steady-state distortions are not too large, the value of λ implied by this formula is positive, as assumed in the earlier paper.

[31] Here we assume, as in our discussion above, that steady-state output is inefficiently low, so that $\overline{\Gamma}_1 > 0$.

endogenous variables, such that for any disturbance process satisfying the bound there exists a unique REE in which the endogenous variables remain always within the neighborhood).

If there exists a policy commitment $\bar{c}$ satisfying these assumptions, then we can show that, up to a log-linear approximation, $\bar{c}$ represents a robustly optimal policy commitment that implements the upper-bound solution defined above.

More precisely, our key result can be stated as follows.

Proposition 4. *Suppose there exists a policy commitment $\bar{c}$ that satisfies Assumptions1 and 2. Then it is possible to define the bound $\|\xi\| > 0$, the neighborhood $\mathcal{N}$; the class of belief distortions $\mathcal{M}$; and a particular policy commitment $c^* \in \mathcal{C}$, where $\mathcal{C}$ is the class of rules specified in Definition 3, under which the policymaker commits to ensure that certain relations $c_t^*(\cdot) = 0$ hold for all t; such that*:

1. *for each t, the function $c_t^*(\cdot)$ is equal to the function $\bar{c}_t(\cdot)$, to a log-linear approximation (i.e., the log-linearizations of policies $\bar{c}$ and c^* are identical);*
2. *in the case of any disturbance process satisfying the bound, and any outcome function $O(m,c)$, defined for all belief distortions $m \in \mathcal{M}$ and all policy commitments $c \in \mathcal{C}$, and associating to any pair (m, c) a DEE in which the endogenous variables remain forever in the neighborhood $\mathcal{N}$, the policy commitment c^* solves the robustly optimal policy problem*

$$\max_{c \in \mathcal{C}} \min_{m \in \mathcal{M}} W(O(m,c)) + \theta V(m); \tag{69}$$

and

3. *the belief distortions m^* that solve the inner problem in (69) are identical to the worst-case belief distortions m^* associated with the upper-bound solution; and the dynamics of the endogenous variables given by the outcome function $O(m^*,c^*)$ are identical to the dynamics under the upper-bound solution.*

Hence by choosing a policy commitment that (to a log-linear approximation) corresponds to a policy $\bar{c}$ that satisfies the conditions stated in the proposition, monetary policy can implement the upper-bound outcome regardless of the assumed outcome function $O(\cdot,\cdot)$, as long as the outcome function selects only equilibria in the neighborhood of the optimal steady state. The proof of Proposition 4 is given in Appendix A.3.

Note that our assumption that the policy commitment $\bar{c}$ (and hence similarly the robustly optimal policy commitment c^*) is expressed by a sequence of backward-looking functions of the path of the particular variables mentioned in Assumption 1 is not intended to imply that this is the *only* coherent formulation of a policy commitment, nor that *only* rules of this form can possibly be robustly optimal policies. We simply have established that it is not *necessary* for the policy commitment to be of some more complex form – for example, it is not necessary either for the policy commitment to refer explicitly to the evolution of the belief distortions $\{m_{t+1}\}$ or to private-sector forecasts of any variables – in order for a robustly optimal policy commitment to exist. In fact, we show below that there are several ways in which once can find robustly optimal policy commitments that have the form assumed in the proposition. Commitments from this simple class have the advantage that the policymaker does not have to commit to a specific empirical measure of private-sector belief distortions when stating its policy commitment.

The corollaries below present a number of specific policy commitments that satisfy the conditions stated in Proposition 4. Among other things, these examples verify that there do exist policy rules satisfying Assumptions 1 and 2.

Corollary 1. *If monetary policy commits to implement the state-contingent inflation sequence of the upper-bound solution (as implied by the solution to Eqs.* (57)–(59)), *then the upper bound is the locally unique outcome of the robust monetary policy problem* (69).

For the commitment considered in the previous corollary, condition 1 of Assumption 2 holds by assumption; and as is easily shown, the commitment also implies a locally determinate outcome under rational expectations, so that the second condition of Assumption 2 holds as well.

The following result shows that monetary policy can alternatively implement the upper bound outcome by committing to a Taylor rule.[32]

Corollary 2. *Suppose monetary policy commits to follow a Taylor rule of the form*

$$1 + i_t = (1 + i_t^*) \left(\frac{\Pi_t}{\Pi_t^*} \right)^{\phi_\Pi} \left(\frac{Y_t}{Y_t^*} \right)^{\phi_Y}, \tag{70}$$

where (i_t^*, Π_t^*, Y_t^*) denotes the evolution of the interest rate, inflation and output in the upper-bound solution. If the coefficients (ϕ_Π, ϕ_Y) satisfy the local determinacy conditions under private-sector rational expectations, then the upper-bound solution is the locally unique outcome of the robust monetary policy problem (69).

[32] Note that conditions 1 and 2 of Assumption 2 are guaranteed to hold by the hypotheses of Corollary 2.

Finally, monetary policy could implement the upper-bound outcome instead by committing to a targeting rule. In this case somewhat more stringent conditions apply.

Corollary 3. *Suppose steady-state output falls short of its efficient level* ($\overline{Y} < \overline{Y}^e$), *and the steady state output distortions are sufficiently small. If monetary policy commits to insure that the target criterion*

$$\xi_\pi \pi_t + \lambda_x (x_t - x_{t-1}) + \xi_m \lambda_m (\pi_t - E_{t-1}[\pi_t]) = 0 \tag{71}$$

holds each period, then the upper-bound solution is the locally unique outcome of the robust monetary policy problem (69).

Condition 1 of Assumption 2 holds because the targeting rule (71) is implied by the log-linearized upper-bound dynamics (58) and (59). Appendix A.3 shows that condition 2 of Assumption 2 also holds, provided that the additional conditions stated in the corollary are satisfied.

This corollary generalizes the RE analysis in Benigno and Woodford (2005) and Giannoni and Woodford (2010), which showed that an optimal policy commitment could be implemented (up to a log-linear approximation) by committing to a linear target criterion. The target criterion (71) differs from the one shown to be optimal in the RE analysis by the presence of the inflation surprise term, multiplied by the coefficient $\xi_m \lambda_m > 0$. The presence of the additional term requires greater resistance to surprise movements in inflation than is implied by the RE target criterion. In order for a surprise increase in inflation to conform to the criterion it must coincide with a surprise decrease in the output gap that is $(\xi_\pi + \xi_m \lambda_m)/\lambda_x$ times as large, whereas the required output gap decline would be only ξ_π/λ_x times as large under the RE target criterion.

To sum up, this section has shown that monetary policy can implement the upper-bound solution as the locally unique outcome of the robustly optimal policy problem by making an appropriate policy commitment. Importantly, the required policy commitments do not need to make explicit reference to private-sector belief distortions, and thus are not fundamentally more difficult to explain to the public than policy commitments that would be desirable under the assumption of private-sector rational expectations.

7. Maximally robust optimal policy

The previous sections were concerned with monetary policy rules that implement the best possible level of policymaker objective under worst-case private-sector beliefs. We now ask whether one can find monetary policy rules that are even more robust, in the sense that they perform better than the robust policy considered above in the case of some possible private-sector beliefs other than the worst-case beliefs,[33] while doing equally well in the case of the worst-case beliefs.

The best that monetary policy can do in response to general belief distortions is to bring about the highest-welfare equilibrium consistent with the given belief distortions, regardless of what those belief distortions may be. This is the outcome that would result if, purely hypothetically, the private sector had to commit to particular belief distortions *before* the policymaker's choice of its policy commitment, and the policymaker could observe those distortions before making its decision. Again, this defines a problem that can be formulated and solved without reference to any particular class of policy commitments – it is simply necessary to optimize over the set of paths for the endogenous variables that constitute a DEE under the given belief distortions – and again this provides an upper bound for what can conceivably be achieved by any policy. If a policy commitment can then be found that achieves this upper bound, it would necessarily be a maximally robust optimal policy.

Under the present, stronger criterion for robustness, it is less obvious that we should expect that the upper bound can be attained; certainly a much more complex type of policy commitment will have to be contemplated if this is to be possible. Nonetheless, here we restrict our discussion to a derivation of the state-contingent evolution corresponding to this upper bound. The following proposition locally characterizes the best response dynamics for output and inflation for a general belief distortion process.[34]

Proposition 5. *If initial price dispersion* Δ_{-1} *is small* (*of order* $\mathcal{O}(\|\xi\|^2)$) *and the initial precommitments such that* $\Gamma_{1,0} = -\Gamma_{2,0} > 0$, *then Eqs.* (42)–(44) *and* (47)–(51) *imply up to first order that the best response dynamics of output and inflation for any given process of belief distortions satisfy*

$$\pi_t = \kappa x_t + \beta E_t \pi_{t+1} + u_t, \tag{72}$$

$$0 = \xi_\pi \pi_t + \lambda_x (x_t - x_{t-1}) + \xi_m \hat{m}_t, \tag{73}$$

where again $\hat{m}_t \equiv \log m_t$, *and the constants* $(\kappa > 0, \xi_\pi, \xi_m, \lambda_x, \lambda_m)$ *satisfy the conditions stated in Proposition* 2.

For the particular case that private sector belief distortions are given by worst-case belief distortions, the previous result reduces to the one given in Proposition 2. For a general process of belief distortions and if the evolution of mark-up

[33] For example, Benigno and Paciello (2010) hypothesize a particular kind of private-sector belief distortions, resulting from a concern for robustness on the part of the public, and suppose that the policymaker should be able to predict this kind of concern on the part of the public. The maximally robust optimal policy characterized in this section would also represent an optimal policy under that hypothesis.

[34] The proof of Proposition 5 follows directly from the steps of the proof of Proposition 2 up to Eq. (90).

shocks is of the autoregressive form (61), Proposition 5 implies that the best response dynamics are given (to first order) by the following recursion

$$\begin{pmatrix} x_t \\ \pi_t \end{pmatrix} = \begin{pmatrix} e_2 \\ \dfrac{\lambda_x(1-e_2)}{\xi_\pi} \end{pmatrix} x_{t-1} + \begin{pmatrix} -\dfrac{\xi_\pi}{\beta\lambda_x(e_1-\rho)} \\ \dfrac{1}{\beta(e_1-\rho)} \end{pmatrix} u_t + \begin{pmatrix} -1\dfrac{\xi_m}{e_1\beta\lambda_x} \\ \dfrac{1-e_1\beta}{e_1\beta}\dfrac{\xi_m}{\xi_\pi} \end{pmatrix} \hat{m}_t, \tag{74}$$

where $e_1 > \beta^{-1}$ and $e_2 \in (0,1)$. This is shown in Appendix A.4. Since $e_1\beta > 1$, the best response dynamics imply that monetary policy optimally reduces inflation in states to which private agents assign higher than objective likelihood ($\hat{m}_t > 0$) and increases it in states whose likelihood private agents underpredict ($\hat{m}_t < 0$).

We also have the following result, which is proven in Appendix A.4:

Proposition 6. *Suppose that monetary policy commits to implement the state-contingent best-response dynamics for inflation, defined in* (74). *Then the worst-case belief distortions are the distortions* $\{\hat{m}_t\}$ *defined in* (59), *or more precisely are given by solution of the system consisting of Eqs.* (59) *and* (74). *The solution to the latter system of equations is again the upper-bound dynamics characterized in Section* 5, *and the associated worst-case value of the augmented objective* (2) *is the same as under the upper-bound solution.*

This shows that committing to the best-response dynamics for inflation, instead of the upper-bound process for inflation, as a monetary policy commitment comes at no cost, if the criterion used to evaluate alternative policy commitments is simply the value of the augmented objective under worst-case beliefs. However, under other types of belief distortions than the ones that would be *worst* for the policymaker, the best-response commitment will in general deliver a higher value for the policymaker's objective than that guaranteed by the upper-bound dynamics for inflation.

The example of a maximally robust policy commitment just given requires the policy commitment to make explicit reference to the magnitude of private-sector belief distortions. While we have no proof, it seems likely that a maximally robust policy will generally refer to a larger set of state variables than the ones allowed by the class $\mathcal{C}$ of policy commitments defined earlier.

8. Conclusions

We have shown how it is possible to analyze optimal monetary stabilization policy, taking into account the possibility that private-sector expectations may not be precisely model-consistent. Moreover, we have shown how to characterize robustly optimal policy without restricting consideration a priori to a particular parametric family of candidate policy rules. Among the policy commitments that we have shown achieves the highest possible lower bound for welfare, for any private-sector beliefs that are not too different from the policymaker's beliefs, is a commitment to a particular target criterion, that maintains a linear relationship between the paths of inflation and of a suitably defined output gap. This optimal target criterion is similar to the one derived by Benigno and Woodford (2005) and Giannoni and Woodford (2010) for the case of rational expectations, except that it no longer refers solely to variations in inflation, regardless of the extent to which these may be anticipated in advance. The additional inflation-surprise term in the modified target criterion implies that a concern for robustness requires greater resistance to surprise increases in inflation than would be considered optimal if one could count on the private sector to have rational expectations.

Among the implications of this change in the target criterion is the fact that an optimal policy commitment no longer implies complete stationarity of the long-run price level. However, we do not feel that this result does much to weaken the case for the desirability of a (suitably flexible) price-level target. By comparison with the type of forward-looking inflation targets actually adopted by inflation-targeting central banks – under which temporary departures of the inflation rate from its long-run target are allowed to persist for a time and are certainly never reversed – a price-level target, which would require temporary departures from the price-level target path to eventually be reversed, would still be closer to the policy recommended by our analysis, in which complete reversal should always occur.

Our specific conclusions depend, of course, on a specific conception of which kinds of departures from model-consistent expectations should be regarded as most plausible. We have proposed a non-parametric specification of the possible belief distortions that is intended to be fairly flexible, but we are well aware that in some ways our specification remains fairly restrictive. In particular, our assumption that the only belief distortions that are contemplated in the robust policy analysis are ones that are absolutely continuous with respect to the policy analyst's own probability measure – a restriction that is necessary in order for our relative entropy measure of the "size" of belief distortions to be defined – is hardly an innocuous one. We are concerned that this assumption may have an important effect on our results. It implies that a determination on the part of the central bank to ensure that a certain relation among variables will hold in all states of the world is sufficient to ensure that the private sector cannot doubt that it will hold in all states of the world; and such an assumption may well still exaggerate the extent to which central bank policy commitments can shape private-sector expectations, even if not to the extent that an assumption of fully model-consistent expectations would. This may lead us to exaggerate the value of a policy commitment to inflation stabilization. An extension of our analysis to allow for alternative definitions of "near-rational expectations" would accordingly be of great value in further clarifying the nature of a robust approach to the conduct of monetary policy.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org.10.1016/j.jmoneco.2012.05.003.

References

Adam, K., 2004. On the relation between Bayesian and robust decision making. Journal of Economic Dynamics and Control 28, 2105–2117.
Benigno, P., Paciello, L., 2010. Monetary Policy, Doubts and Asset Prices. LUISS Guido Carli (Rome) Working Paper.
Benigno, P., Woodford, M., 2005. Inflation stabilization and welfare: the case of a distorted steady state. Journal of the European Economic Association 3, 1185–1236.
Calvo, G.A., 1983. Staggered contracts in a utility-maximizing framework. Journal of Monetary Economics 12, 383–398.
Clarida, R., Galí, J., Gertler, M., 1999. The science of monetary policy: evidence and some theory. Journal of Economic Literature 37, 1661–1707.
Giannoni, M., Woodford, M., 2010. Optimal target criteria for stabilization policy. NBER Working Paper No. 15757.
Hansen, L.P., Sargent, T.J., 2005. Robust estimation and control under commitment. Journal of Economic Theory 124, 258–301.
Hansen, L.P., Sargent, T.J., 2008. Robustness. Princeton University Press, Princeton.
Hansen, L.P., Sargent, T.J., 2011. Wanting robustness in macroeconomics. In: Friedman, B.M., Woodford, M. (Eds.), Handbook of Monetary Economics, vol. 3B. , Elsevier, Amsterdam.
Hansen, L.P., Sargent, T.J. Two Views of a Robust Ramsey Planner. University of Chicago, January, unpublished.
Woodford, M., 2010. Robustly optimal monetary policy with near-rational expectations. American Economic Review 100, 274–303.
Woodford, M., 2011. Optimal monetary stabilization policy. In: Friedman, B.M., Woodford, M. (Eds.), Handbook of Monetary Economics, vol. 3B, Elsevier, Amsterdam
Yun, T., 1996. Nominal price rigidity, money supply endogeneity, and business cycles. Journal of Monetary Economics 27, 345–370.