



**THE PINHAS SAPIR CENTER FOR DEVELOPMENT
TEL AVIV UNIVERSITY**

“Managing Non-Rational Expectations: A “Monetary Policy” Example”

Ran Spieglerⁱ

Discussion Paper No.3-16

March 2016

Financial support by the The Pinhas Sapir Center for Development at Tel Aviv University is gratefully acknowledged.

I thank Kfir Eliaz, Rafaella Giacomini, Heidi Thysen, Michal Woodford and Vasiliki Skreta for helpful discussions.

ⁱRan Spiegler – The Eitan Berglas School of Economics, Tel Aviv University. University College of London and CFM.
Email: rani@post.tau.ac.il

Abstract

I study a stylized static model of optimal monetary policy in the tradition of Barro and Gordon (1983), which relaxes the conventional assumption that the private sector has rational expectations. Building on the literature on Bayesian networks, I assume that the private sector forms its inflation forecast by fitting a misspecified causal model - formalized as a direct acyclic graph (DAG) over relevant variables - to the objective steady-state distribution. I characterize the central bank's ex-ante optimal contingent policy for various private-sector DAGs. The classical impossibility of using monetary policy to systematically enhance real activity persists for a large class of DAGs. In these cases, the central bank's optimal policy may display excessive rigidity relative to the rational-expectations benchmark.

1 Introduction

Many real-life interactions can be described as a principal-agent problem, in which the principal's ability to achieve his objectives depends on whether the agent correctly anticipates the principal's actions or their consequences. In some situations, the principal wishes to surprise the agent. For instance, success of a police crackdown on a drug-trafficking operation hinges on its unpredictability. Likewise, the immediate effect of a pay rise on worker morale is intuitively larger when it comes as a surprise. In other situations, the principal would like the agent to hold correct expectations. For example, when a company's management adapts its marketing strategy to changes in consumer demand, it would be better served if sales and service staff could anticipate the adaptation in advance, for the sake of swift implementation of the new strategy.

Monetary theory offers a prominent example of this general idea. In a well-known class of models, originated by Kydland and Prescott (1977) and Barro and Gordon (1983), the central bank controls a policy variable that affects inflation. The private sector forms an inflation forecast, possibly after observing some signal regarding the central bank's decision. Private-sector expectations are relevant because real output (or unemployment) is determined by an "expectations-augmented" Phillips curve, such that the real effect of inflation is at least partly offset when inflation is anticipated.

Thus, to the extent that the central bank wishes to maximize expected output, it would like to set inflation systematically above private-sector expectations. And to the extent that the central bank wishes to minimize the variance of output fluctuations, it would like to avoid inflationary surprises. It follows that monetary policy involves "*expectations management*". To quote Woodford (2003, p. 15):

"...successful monetary policy is not so much a matter of effective control of overnight interest rates as it is of shaping market expectations of the way in which interest rates, inflation and income are likely to evolve..."

Conventional models constrain the central bank's ability to manage expectations by assuming that the private sector has "rational expectations" - i.e., it fully understands the statistical regularities in its environment, and thus forms an unbiased inflation forecast conditional on its information. This constraint immediately implies that the central bank is unable to systematically fool the private sector. The literature (e.g., Athey et al. (2005)) has thus mainly focused on the private sector's *information*: Can it monitor the central bank's moves - and if not, can the central bank commit ex-ante to a contingent strategy? Does it observe the macro variables that guide the central bank's policy?

In this paper, I revisit the textbook problem of optimal monetary policy and relax the assumption that the private sector has rational expectations. I adapt a simple reformulation of the Barro-Gordon model due to Sargent (2001), and examine a static environment in which a central bank commits ex-ante to a contingent strategy that assigns a (possibly random) action to every realization of some state of Nature that defines an "ideal" inflation target. Actual inflation is a stochastic function of the central bank's action. The private sector forms its inflation forecast after perfectly observing the central bank's move. Thus, if the private sector had rational expectations, it would be able to form an optimal forecast. Real output is a stochastic function of inflation and private-sector inflationary expectations, given by a linear Phillips Curve that allows both anticipated and surprise inflation to have real effects. The central bank's objective function involves three motives: enhancing expected output, minimizing output variance, and minimizing the expected square deviation of inflation from the ideal target.

I depart from the conventional literature by relaxing the assumption that the private sector has rational expectations. Of course, one could model non-rational expectations in various ways. My approach is based on the following simple idea: *the private sector derives its expectations by fitting a misspecified subjective causal model to the steady-state probability distribution over relevant variables* (which in turn is induced by the exogenous stochastic processes and the central bank's strategy). This idea seems relevant in the present context for at least two reasons. First, private-sector agents hold

intuitive, qualitative theories about the interconnection between real and monetary variables. Such theories can sometimes be viewed as statements about causal relations. Indeed, Hoover (2001) describes historical controversies in macroeconomics in such terms. Second, key private-sector actors (banks, financial-market speculators) regularly employ statistical models to form macroeconomic forecasts. While the exact specification of these models may be tweaked from time to time in order to get good empirical fit, their basic underlying causality assumptions are more likely to remain constant during times of relative stability. (For a study of how macroeconomic forecasters rely on models, see Giacomini et al. (2015).)

To formalize the notion that expectations are formed by fitting a misspecified causal model, I employ a recent modeling framework (Spiegler (2015a)), which in turn builds on the Statistics and Artificial-Intelligence literature on *Bayesian networks* (Cowell et al. (1999), Pearl (2009)). The following example illustrates the modeling approach and its possible implications.

An example: Belief in a “classical” causal model

Suppose that the ideal inflation target is fixed, and consider three macro variables: the central bank’s action a , inflation π and real output y . The private-sector causal model is represented by a *directed acyclic graph* (DAG), denoted R , which is defined over the three variables:

$$a \rightarrow \pi \leftarrow y \tag{1}$$

A direct link between two nodes signifies a perceived direct causal link between the variables they represent. Thus, the DAG R captures a causal model according to which inflation is a consequence of two independent causes: output and the central bank’s action. The causal model is entirely non-parametric - it does not assume anything regarding the sign or magnitude of causal relations - it merely postulates their existence and direction.

The causal model R is false because it perceives output to be exogenous and thus independent of monetary policy, whereas according to the true process it is a consequence of the central bank’s action via the Phillips Curve. Thus, the private sector subscribes to a “classical worldview” that

postulates the absolute neutrality of monetary policy, whereas the true model given by Sargent (2001) allows for non-neutrality. Another way of expressing this disagreement is that R postulates that output causes inflation, whereas according to the true model, causation runs in the opposite direction. In other words, the private sector’s subjective model exhibits *reverse causality*.

How does the private sector employ its causal model to forecast inflation? It simply *fits* the model to the true steady-state joint distribution p over a, π, y . If p were indeed consistent with R , $p(a, \pi, y)$ could be written as

$$p_R(a, \pi, y) = p(a)p(y)p(\pi | a, y) \tag{2}$$

The formula $p_R(a, \pi, y)$ describes the private sector’s subjective belief as a function of the true steady-state distribution p . It is an example of a “*Bayesian network factorization formula*” - it factorizes the steady-state distribution p into a product of conditional-probability terms, *as if p were consistent with R* . This is how I formalize the notion that the DM “fits a subjective causal model to the steady-state distribution”. Note that because the causal model is entirely non-parametric, the private sector is always able to perfectly fit it to any objective distribution.

The subjective belief p_R systematically distorts the true correlation structure of the steady-state distribution p . The distortion arises because the private sector perceives statistical regularities through the prism of an incorrect causal model. Specifically, the private sector’s inflation forecast after observing the central bank’s action a is

$$E_R(\pi | a) = \sum_{\pi} p_R(\pi | a)\pi = \sum_{\pi} \left(\sum_y p(y)p(\pi | a, y) \right) \pi$$

This is in general different from the “rational” inflation forecast

$$E(\pi | a) = \sum_{\pi} p(\pi | a)\pi = \sum_{\pi} \left(\sum_y p(y | a)p(\pi | a, y) \right) \pi$$

The discrepancy arises because $p_R(\pi | a)$ involves an implicit expectation

over y *without* conditioning on a . Note that if the private sector could - or felt the need to - test its causal model against historical data, it would discover the discrepancy between $p_R(\pi | a)$ and $p(\pi | a)$, thus refuting the model. I assume that no such “test for model misspecification” occurs - see Spiegler (2015a,b) for a discussion of this feature of the modeling approach.

The question is how the private sector’s “non-rational” inflation forecast affects the central bank’s considerations. It turns out that there are specifications of the exogenous processes (the transmission from a to π , the distribution of the error term in the Phillips Curve) for which the central bank can create systematic underestimation of inflation by the private sector, thus enhancing expected output. The optimal policy involves *randomizing* over a . Thus, somewhat ironically, the central bank’s ability to systematically fool the private sector arises because it believes that inflation is a consequence of real output, rather than the other way around. \square

Although the example’s punchline is that the central bank can sometimes systematically fool the private sector, this is by no means the rule. The main characterization results, given in Section 4, show that for a large subclass of private-sector DAGs R , the inflation forecast $E_R(\pi | a)$ is unbiased on average, regardless of how we specify the exogenous processes. The result is stronger under a conventional linear-normal specification of these processes: when the inflation target is constant, systematic output enhancement is impossible for *any* R .

This collection of impossibility results is intriguing, considering the heated historical debates over the exploitability of the Phillips relation (see Klammer (1984)). The key assumption behind classical non-exploitability results (Lucas (1972), Sargent and Wallace (1975)) was allegedly the rationality of private-sector expectations. However, according to this paper, a considerably milder assumption - namely that the private sector forms its expectations by fitting a (potentially misspecified) causal model to long-run data - reproduces results in a similar vein.

If the private sector’s false causal perceptions do not make it easy for the central bank to enhance expected output, this does *not* mean that they are irrelevant for monetary policy. Indeed, when the central bank cares about

minimizing output variance, the conditions for the persistence of the rational-expectations prediction are much more stringent. The reason is that even when the private sector's inflation forecast is unbiased on average, it can be less responsive to the central bank's action than in the rational-expectations benchmark. For instance, if R postulates that inflation is independent of the central bank's action, the private sector will form its inflation forecast as if it has not observed it. This impels the central bank to adopt a relatively *rigid* policy, in order to curb output fluctuations.

This paper is related to a few works that examine monetary policy when the rational-expectations assumption is relaxed. Evans and Honkapohja (2001) and Woodford (2013) review dynamic models in which agents form various forms of non-rational expectations, and explore implications for monetary policy. See Garcia-Schmidt and Woodford (2015) for a recent exercise in this tradition. Sargent (2001), Cho et al. (2002) and Esponda and Pouzo (2015) study models in which it is the central bank that forms non-rational expectations, whereas the private sector is modeled conventionally.

More broadly, this paper contributes to the literature (reviewed in Spiegler (2015a)) that studies strategic interaction among agents who base their decisions on misspecified subjective models. Within this literature, Piccione and Rubinstein (2003) shares the "expectations management" aspect of the present paper. In their model, the principal is a seller who commits to a deterministic temporal sequence of prices, taking into account that consumers can only perceive statistical patterns that allow the price at any period t to be a function of price realizations at periods $t - 1, \dots, t - k$, where k is a constant that characterizes the consumer. When the value of k is negatively correlated with consumers' willingness to pay, the seller may want to generate a complex price sequence as a price-discrimination device.

Finally, as the scare quotes in the title suggest, the monetary-policy terms in which I chose to cast my story are quite arbitrary. The model is so stylized that it could just as well accommodate other principal-agent situations - referred to in the opening paragraph - in which expectations management is part of the principal's problem. The monetary-policy terminology is justified because precisely such stylized models served a decisive role in macro-

economists’ debates over the role of expectations in monetary policy. The monetary-theory literature thus provides an extremely rich background for the current exercise. Hopefully, this exercise will stimulate further research on the role of misspecified subjective causal models in macroeconomics.

2 The Model

A central bank (CB henceforth) chooses an action a after observing the realization of a state of Nature θ . The rate of inflation π is a probabilistic function of the action a . The private sector (PS henceforth) observes CB’s move and forms an inflationary expectation e . I assume that PS does *not* observe the realization of θ . Real output y is a probabilistic function of π and e that obeys the following “expectations-augmented Phillips Curve”:

$$y = \gamma\pi - e + \eta \tag{3}$$

where $\gamma \geq 1$ is a constant and η is an independent random variable with zero mean. When $\gamma = 1$, anticipated inflation has no real effect.

A policy for CB is a function that assigns a probability distribution over a to every θ . I assume that CB is able to *commit* to a policy ex-ante. I use p to denote the joint distribution over (θ, a, π, e, y) . By the above description, p can be factorized into a collection of marginal and conditional probabilities:

$$p(\theta)p(a | \theta)p(\pi | a)p(e | a)p(y | \pi, e) \tag{4}$$

where $p(\theta)$, $p(\pi | a)$ and $p(y | \pi, e)$ represent the exogenous components (where the latter component is given by (3)); $p(a | \theta)$ represents CB’s strategy; and $p(e | a)$ represents PS’s belief formation.

When PS has rational expectations, $p(e | a)$ assigns probability one to $E(\pi | a) = \sum_{\pi} p(\pi | a)\pi$. Other models of PS’s belief formation will give rise to different restrictions on how $p(e | a)$ relates to the other components of p .

Note that because π is assumed to be independent of θ conditional on a , the assumption that PS does not observe θ has no importance when PS has

rational expectations - i.e., $E(\pi | a) = E(\pi | \theta, a)$. The assumption remains inessential when I relax the rational-expectations assumption; its role is to simplify notation and the statement of general results.

2.1 Causal Models and PS Expectations

I now wish to introduce the idea - originally formulated in Spiegler (2015a) - that PS forms non-rational expectations by fitting a misspecified causal model to long-run data. This will require basic concepts from the literature on Bayesian networks. The following exposition is standard (e.g., see Cowell et al. (1999) and Pearl (2009)), with a few minor adjustments that serve the current purposes.

First, denote $x = (\theta, a, \pi, e, y)$, and impose an arbitrary enumeration over the five variables, such that $x = (x_1, x_2, x_3, x_4, x_5)$. Let $N = \{1, 2, 3, 4, 5\}$ denote the set of the variables' labels or indices. Define a *directed acyclic graph* (DAG) (N, R) , where N is the set of nodes and R is the set of directed links. (A directed graph is acyclic if it does not contain a directed path from a node to itself.) I use jRi or $j \rightarrow i$ interchangeably to denote a directed link from j into i . Observe that the binary relation R is asymmetric and acyclic. For every $i \in N$, slightly abusing notation, let $R(i) = \{j \in N \mid jRi\}$ be the set of "direct parents" of node i . From now on, I will identify the DAG with R . Let \tilde{R} be the *skeleton* (or undirected version) of R - i.e., $i\tilde{R}j$ if and only if iRj or jRi . A subset $M \subseteq N$ is a *clique* in R if $i\tilde{R}j$ for every $i, j \in M$. A clique M is *ancestral* if $R(i) \subset M$ for every $i \in M$.

PS is characterized by a DAG R . For any objective joint probability distribution p over x , PS's subjective belief is

$$p_R(x) = \prod_{i \in N} p(x_i \mid x_{R(i)}) \quad (5)$$

A probability distribution p is *consistent* with R if $p_R(x) \equiv p(x)$.

Following Peal (2009), I interpret R as a *causal model*. The link $j \rightarrow i$ means that PS regards the variable x_j to be an immediate cause of the variable x_i . While PS presupposes the existence of this causal effect, it has

no preconception regarding its sign or magnitude. In particular, this effect could be measured to be null. The formula (5) represents PS's attempt to fit the causal model to objective data generated by the true process.

Note that R is entirely non-parametric: although the "true model" has parametric structure given by the Phillips Curve (3), PS does not impose any parametric restriction and fits its non-parametric causal model R to the data generated by the true model. For a concrete image to match this description, think of an analyst who tries to fit data with a system of structural equations. The analyst holds the collection of R.H.S variables fixed, but tweaks the exact functional form, until he gets good fit.

PS's subjective distribution over π conditional on the observation of a is

$$p_R(\pi | a) = \frac{p_R(a, \pi)}{p_R(a)}$$

where¹

$$\begin{aligned} p_R(a, \pi) &= \sum_{\theta, e, y} p_R(\theta, a, \pi, e, y) \\ p_R(a) &= \sum_{\pi'} p_R(a, \pi') \end{aligned}$$

I impose the restriction that for every a , $p(e | a)$ assigns probability one to

$$e_R(a) = \sum_{\pi} p_R(\pi | a)\pi \tag{6}$$

This restriction captures the interpretation of e as PS's inflation forecast: it is expected inflation according to PS's beliefs, conditional on the observed realization of a .

Because the true joint distribution p can be written as the factorization

¹Throughout the paper, I use simple summations rather than integration, for notational clarity.

formula (4), p is consistent with the following “true DAG” R^* - i.e. $p = p_{R^*}$:

$$\begin{array}{ccccc} \theta & \rightarrow & a & \rightarrow & \pi \\ & & \downarrow & & \downarrow \\ & & e & \rightarrow & y \end{array}$$

When $R = R^*$ (or when R adds links to R^*), (6) reduces to rational expectations - i.e., $e_R(a) = E(\pi | a)$ for every objective distribution p .

The true DAG R^* reflects the assumption that π is purely a (stochastic) function of CB’s policy, and that y is independent of a conditional on π and e . This is a feature of Sargent’s (2001) model, which sets it apart from the model of Barro and Gordon (1983), where π and y are determined simultaneously. I employ Sargent’s specification for two reasons. First, it is somewhat easier to work with. Second, because Sargent studied a “dual” problem in which PS (CB) has rational (non-rational) expectations, it may be beneficial to make my formulation as close as possible to his. I discuss alternative specifications of the true process in Section 5.

From now on, I impose the following restriction on PS’s DAG R - it perceives a to be the only immediate cause of e - i.e., abusing notation, $R(e) = \{a\}$. This property also holds under the true DAG R^* . Thus, PS understands the true immediate causes of its own expectations. The justification for this assumption is that PS actively conditions its expectations on the observed realization of a , and on no other variable - therefore, it would be odd if its subjective causal model failed to acknowledge this feature. This assumption has the following useful implications.

Lemma 1 *Suppose that $R(e) = \{a\}$, and restrict the domain of objective joint distributions p to be those for which $p(e | a)$ assigns probability one to $e_R(a)$. Then:*

- (i) *There is a DAG R' that omits the node e altogether, such that $p_{R'}$ coincides with p_R over all objective distributions in the restricted domain. In particular, R' establishes a link from a to any node that receives a link from e in R .*
- (ii) *Fix the exogenous components of p , $(p(\theta))_\theta$ and $(p(y | \pi, e))_{\pi, e, y}$, and fix*

CB's strategy $(p(a | \theta))_{\theta, a}$. Then, p is uniquely determined, such that $e_R(a)$ is uniquely determined.

Proof. (i) Consider the factorization formula (5). By our restriction on R , it contains the term $p(e | a)$. By assumption, $p(e_R(a) | a) = 1$, and $p(e | a) = 0$ for all $e \neq e_R(a)$. Therefore, we can remove this term from (5) altogether, and plug $e = e_R(a)$ in any other term in (5) that conditions on e - which effectively means that such a term conditions on a . We have thus obtained a DAG representation in which the node e is omitted, and any link from e to some node in R is replaced with a link from a into the same node.

(ii) By assumption, we restrict attention to objective distributions p for which $p(e_R(a) | a) = 1$. For these distributions p , we can modify the factorization formula (4) and write

$$p(\theta, a, \pi, y) = p(\theta)p(a | \theta)p(\pi | a)p(y | \pi, e_R(a))$$

Because e is effectively omitted from R , $p_R(\theta, a, \pi, y)$ is uniquely defined for every p and every θ, a, π, y . ■

The first part of this result means that we can assume w.l.o.g that PS's DAG R is defined over the four nodes θ, a, π, y , omitting the node e altogether. *I will follow this practice from now on.* The second part means that every CB strategy induces a unique, well-defined joint distribution p , such that $e_R(a)$ is uniquely determined.

Equivalent DAGs

In this model, the DAG R represents a mapping from any objective distribution p to PS's subjective belief p_R . Two DAGs can be equivalent in the sense that they represent the same mapping.

Definition 1 *Two DAGs R and Q are **equivalent** if $p_R(x) \equiv p_Q(x)$ for every $p \in \Delta(X)$.*

For instance, the DAGs $1 \rightarrow 2$ and $2 \rightarrow 1$ are equivalent by the basic identity $p(x_1)p(x_2 | x_1) \equiv p(x_2)p(x_1 | x_2)$. Thus, two different causal models

can be indistinguishable in terms of the statistical regularities they induce. In particular, a DAG that involves intuitive causal relations can be equivalent to a DAG that makes little sense as a causal model (e.g., it postulates a causal link between two variables in a direction that contradicts the temporal sequence of their realizations).

The following characterization of equivalent DAGs will be useful in the sequel. A *v-collider* in R is an ordered triple of nodes (i, j, k) such that iRk , jRk , $i \not R j$ and $j \not R i$ (that is, R contains links from i and j into k , yet i and j are not linked to each other). We will say in this case that there is a *v-collider into k* .

Proposition 1 (Verma and Pearl (1991)) *Two DAGs R and Q are equivalent if and only if they have the same skeleton and the same set of v-colliders.*

To illustrate this result, all fully connected DAGs have the same skeleton (every pair of nodes is linked) and an empty set of *v-colliders*, hence they are all equivalent. Indeed, they all induce rational expectations because they reduce (5) to the textbook chain rule. In contrast, the DAGs $1 \rightarrow 2 \rightarrow 3$ and $1 \rightarrow 2 \leftarrow 3$ are not equivalent: although their skeletons are identical, the former DAG has no *v-colliders* whereas $(1, 3, 2)$ is a *v-collider* in the latter DAG.

In the present context, the notion of equivalent DAGs means that PS's causal model may differ from R^* and yet have rational expectations. In particular, recall that by Lemma 1, we can omit the node e from R^* and rewrite it as

$$R^* : \begin{array}{ccccc} \theta & \rightarrow & a & \rightarrow & \pi \\ & & & \searrow & \downarrow \\ & & & & y \end{array}$$

From now on, this specification of R^* will be referred to as the true DAG. Because this DAG contains no *v-colliders*, any R with the same skeleton (or any DAG that contains R) will give rise to rational expectations.

One of the main virtues of this modeling approach is that it enables us to check whether certain marginal or conditional distributions that are

induced by PS's subjective belief do not distort the true distribution, by examining structural properties of the PS's DAG. In particular, the following two properties will be useful in the sequel.

Remark 1 *Suppose that R contains no directed path from a node i - or from one of its ancestors - into another node j . Then, R perceives x_i and x_j to be unconditionally independent - i.e., $p_R(x_j | x_i) = p_R(x_j)$.*

Remark 2 *Let R be a DAG. Let A be a set of nodes. Then, $p_R(x_A) \equiv p(x_A)$ for all objective distributions p if and only if A is an ancestral clique in some DAG R' which is equivalent to R .*

Remark 1 follows from the more general notion of d -separation (see Pearl (2009), or Appendix B in Spiegler (2015a)). Remark 2 is proved in Spiegler (2015b). In particular, it implies that when a node i is ancestral in R (or in some other DAG which is equivalent to R), PS's subjective marginal distribution over x_i will coincide with the true marginal.

2.2 CB's Problem

Let us first define CB's objective function. I assume that CB evaluates joint distributions p according to the following utility function:

$$U(p) = k_e \cdot E(y) - k_v \cdot Var(y) - k_t \cdot E(\pi - \theta)^2$$

where $k_e, k_v, k_t \geq 0$ are constants. The interpretation is that CB has three motives. First, it wishes to increase expected output. Second, it wishes to reduce the variance of output. Finally, every state of Nature is identified with a distinct "inflation target", and CB is averse to deviations of actual inflation from the target.

Comment: CB's objective function is an extension of mean-variance preferences. In the literature, it is more conventional to assume that CB is an expected-utility maximizer whose vNM function is $u(\theta, \pi, y) = -(y - y^*)^2 -$

$k(\pi - \theta)^2$, where $k > 0$ is a constant and y^* represents an ideal output level, because this specification is a reduced-form of a representative agent's welfare in a larger dynamic model (e.g., see Woodford (2003)). The first term in u mixes the first two motives in CB's objective function U . In particular, when y^* and k are high, the output-enhancement motive is dominant. Both approaches yield essentially the same results. The reason I opt for the mean-variance-style preferences given by U is that they provide a cleaner demonstration of the roles of CB's motives.

We can now state CB's optimization problem. Fix the exogenous distributions $(p(\theta))_\theta$, $(p(\pi | a))_{a,\pi}$ and $(p(y | \pi, e))_{\pi,e,y}$. By Lemma 1, any strategy $(p(a | \theta))_{\theta,a}$ by CB uniquely pins down p . In particular, p satisfies $p(e_R(a) | a) = 1$ for every a . Thus, when CB chooses a strategy, it effectively chooses the joint distribution p . We will assume that *CB chooses its strategy ex-ante to maximize U .*

3 Examples

In this section I illustrate the model with three different specifications that highlight various features.

3.1 Persistence of the Rational-Expectations Policy

The following example shows how the rational-expectations prediction can persist even when PS's subjective causal model is misspecified. Let $\gamma = 1$ and $k_v = 0$ - i.e., anticipated inflation has no real effects, and CB does not care about output variance. Assume that PS's subjective DAG is $R : \theta \rightarrow a \rightarrow y \leftarrow \pi$. According to this causal model, PS realizes that CB may condition its policy on the inflation target, but it regards inflation as an exogenous variable. The following result establishes that under this DAG, CB's optimal policy is the same as when PS has rational expectations.

Proposition 2 *Let $\gamma = 1$, $k_v = 0$ and $R : \theta \rightarrow a \rightarrow y \leftarrow \pi$. Then, it is optimal for CB to assign probability one to $\arg \min_a E(a - \theta)^2$ for every θ .*

Proof. Because there is no directed path from a or (its sole ancestor) θ into π , Remark 1 implies $p_R(\pi | a) = p_R(\pi)$. Moreover, because the node π is ancestral in R , Remark 2 implies $p_R(\pi) = p(\pi)$. It follows that

$$\sum_a p(a) \sum_\pi p_R(\pi | a) \pi = \sum_\pi p(\pi) \pi$$

That is, $e_R(a)$ is equal to ex-ante expected inflation for all a . Since $\gamma = 1$, this means that $E(y) = 0$. Since $k_v = 0$, it follows that CB's objective function collapses to

$$-k_t \sum_\theta p(\theta) E(a - \theta)^2$$

Therefore, for every θ , CB chooses a to minimize $E(a - \theta)^2$. ■

In this example, PS regards inflation as an exogenous variable. As a result, $e_R(a)$ is constant in a and therefore does not necessarily coincide with the true expectation of π conditional on a . In this sense, PS lacks rational expectations. However, PS is not systematically fooled - its inflation forecast is unbiased ex-ante. This means that CB is unable to use monetary policy to systematically raise output. Since CB does not care about output variance, it acts as it would if PS had rational expectations - namely, by minimizing the expected square deviation of inflation from the target. The lesson from the example is thus that PS need not have rational expectations in order for the rational-expectations prediction to persist.

3.2 Creating a Biased Inflation Forecast

This sub-section elaborates on the illustrative example from the Introduction. It demonstrates that CB can sometimes use monetary policy to increase expected output, by causing PS to systematically underestimate inflation. As in the previous sub-section, let $\gamma = 1$ and $k_v = 0$. Normalize $k_e = 1$ and denote $k_t = k$. In addition, assume that $\eta = \theta = 0$ with certainty. That is, the Phillips Curve is deterministic, and the inflation target is fixed. The variables a and π take values in $\{0, 1\}$, where $\pi = 0$ (1) represents low (high) inflation, and $p(\pi = 1 | a) = \beta a$, $\beta \in (0, 1)$. Thus, the action $a = 0$ induces

low inflation with certainty, whereas the action $a = 1$ induces high inflation with probability β .

By our assumption that θ is fixed, this variable can be safely removed from PS's DAG. Let

$$R : a \rightarrow \pi \leftarrow y$$

The interpretation is that PS postulates that output fluctuations are exogenous, whereas inflation fluctuations are the consequence of output fluctuations as well as monetary policy. In relation to the true DAG R^* , R reverses the causal link between inflation and output, and it neglects the effect of expectations on output. Under R , we can write $e_R(a)$ as follows:

$$e_R(a) = \sum_{\pi} \sum_y p(y) p(\pi | a, y) \pi$$

Let us now characterize CB's optimal strategy under this specification.

Proposition 3 *Under the specification of this sub-section, CB's optimal strategy is to play $a = 1$ with probability $\min(0, \frac{1-k}{2})$.*

Proof. Denote $p(a = 1) = \alpha$. Denote $e_R(a) = e(a)$. Because $\pi \in \{0, 1\}$,

$$e(a) = \sum_y p(y) p(\pi = 1 | a, y)$$

First, observe that $p(\pi = 1 | a = 0) = 0$. Therefore, $p(\pi = 1 | a = 0, y) = 0$ for all y , hence $e(0) = 0$. This in turn means that conditional on $a = 0$, $y = 0$ with probability one. It follows that when $\alpha = 0$, $\pi = e = y = 0$ with probability one, hence CB's payoff is zero. From now on, assume $\alpha > 0$.

To calculate $p(\pi = 1 | a = 1, y)$, note that since $\pi \in \{0, 1\}$ and the Phillips Curve is deterministic, there is a perfect correlation between π and y conditional on $a = 1$: $y = 1 - e(1)$ when $\pi = 1$, and $y = -e(1)$ when $\pi = 0$. Therefore, $p(\pi = 1 | a = 1, y = 1 - e(1)) = 1$ and $p(\pi = 1 | a = 1, y = -e(1)) = 0$. It follows that

$$e(1) = p(y = 1 - e(1)) \tag{7}$$

Using this identity, we can characterize the relevant quantities for CB's payoff function that are induced by CB's strategy α : $E(\pi)$ and $\sum_a p(a)e(a)$.

By (7), y gets at most three values: $-e(1)$, 0 and $1 - e(1)$. Two of these values coincide if $e(1) \in \{0, 1\}$. Suppose $e(1) = 0$. Then, by (7), $p(y = 1) = 0$. However, given CB's strategy, $\pi = 1$ with probability $\alpha\beta > 0$, hence $\pi - e(1) = 1 > 0$ with positive probability, which is a contradiction. Now suppose $e(1) = 1$. Then, by (7), $p(y = 0) = 1$, hence CB's payoff is at most zero. It follows that we should be interested only in strategies that induce $e(1) \neq 0, 1$, such that y gets precisely three values. In this case, (7) implies that $e(1) \in (0, 1)$, and $y = 1 - e(1)$ if and only if $\pi = 1$. The latter event occurs with probability $\alpha\beta$. We conclude that the only case we need to check is when $e(1) = \alpha\beta$. Thus, $E(\pi) = \alpha\beta$ and $\sum_a p(a)e(a) = \alpha \cdot \alpha\beta + (1 - \alpha) \cdot 0 = \alpha^2\beta$. CB chooses α to maximize

$$(1 - k)\alpha\beta - \alpha^2\beta$$

which immediately gives the solution. ■

Thus, CB benefits from randomizing between its two available actions. This randomization causes PS to systematically underestimate expected inflation, such that if k is sufficiently small, the positive effect on output of this underestimation outweighs the negative effect due to the deviation from the zero-inflation target.

The intuition behind the result is as follows. When the realization of CB's strategy is $a = 0$, it induces $\pi = 0$ with certainty. In this case, PS's incorrect causal model does not lead to a biased estimate of inflation: $p(\pi = 0 | a = 0; y) = p(\pi | a = 0)$ and therefore $p_R(\pi | a = 0)$ assigns probability one to 0. In contrast, calculating PS's inflationary expectations conditional on $a = 1$ involves aggregating over all values of y , such that effectively $E_R(\pi | a = 1) = E_R(\pi)$. Thus, when CB plays $a = 0$, PS correctly updates its belief, whereas when CB plays $a = 1$, PS forms its inflationary expectations as if it has not observed CB's move. As a result, PS effectively "double-counts" the episodes in which CB plays $a = 0$. This leads to systematic underestimation of expected inflation, which CB can exploit to increase expected output.

Finally, note that β is completely irrelevant for CB's strategy, due to the linearity of $E_R(\pi \mid a = 1)$ in β .

3.3 Rigid Policy in a Linear-Normal Specification

In this sub-section, I examine a popular specification of the underlying true process. First, assume that in the Phillips Curve (3), $\eta \sim N(0, \sigma_\eta^2)$. The dependence of inflation on CB's action is given by $\pi = a + \varepsilon$, where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$, and ε and η are mutually independent. I do not impose any structure on $p(\theta)$. Throughout this section, I use μ_z to denote the ex-ante expected value of any variable z according to p .

While the previous two sub-sections focused on the output-enhancing motive, in this sub-section I assume this motive away, and focus on the trade-off between the output-variance and targeting motives. Thus, let $k_e = 0$, normalize $k_v = 1$ and denote $k_t = k$. The CB's objective is thus to minimize

$$\text{Var}(y) + k \cdot E(\pi - \theta)^2$$

Suppose PS has rational expectations. Then, its inflation forecast conditional on a is equal to a , since $\pi = a + \varepsilon$ and ε is an independent variable with mean zero. Therefore, we can write

$$y = (\gamma - 1)a + \gamma\varepsilon + \eta$$

and since ε and η are independent variables with mean zero, we can ignore them in the calculation of the objective function, which is reduced to

$$(\gamma - 1)^2 E[(a - \mu_a)]^2 + kE[a - \mu_\theta]^2$$

Solving this problem is standard. The strategy that minimizes this objective function is as follows. For every θ , it would play $a^*(\theta)$ with certainty, where

$$a^*(\theta) = \frac{k}{(\gamma - 1)^2 + k} \theta + \frac{(\gamma - 1)^2}{(\gamma - 1)^2 + k} \mu_\theta$$

Note that this solution does not rely on the normality assumption - it only requires ε and η to be independent zero-mean random variables.

The optimal policy under rational expectations exhibits some rigidity: it is a weighted average of the realized inflation target θ and the ex-ante average target μ_θ . A higher weight on the latter corresponds to a policy that is less responsive to fluctuations in the state of Nature. As the parameter k (which captures the importance of the targeting motive) increases, CB will more closely track the target. Also, as γ approaches 1 - such that anticipated inflation matters less for output - CB's policy will involve closer targeting.

Now assume that PS's subjective DAG is

$$R : \theta \rightarrow a \rightarrow \pi \leftarrow y$$

Thus, as in the previous sub-section, PS postulates that output fluctuations are exogenous. It correctly perceives the causal chain $\theta \rightarrow a \rightarrow \pi$, but it inverts the direction of causation between π and y , and it fails to acknowledge the effect of a on y via PS's inflationary expectations.

Proposition 4 *Let $k_e = 0$, $k_v = 1$, $k_t = k$; $\pi = a + \varepsilon$, where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$; and $\eta \sim N(0, \sigma_\eta^2)$. Then, CB's optimal policy given R is to play*

$$a^{**}(\theta) = \frac{k}{\lambda(\gamma - 1)^2 + k} \theta + \frac{\lambda(\gamma - 1)^2}{\lambda(\gamma - 1)^2 + k} \mu_\theta$$

for every θ , where

$$\lambda = \left(\frac{\gamma^2 \sigma_\varepsilon^2 + \sigma_\eta^2}{\gamma(\gamma - 1) \sigma_\varepsilon^2 + \sigma_\eta^2} \right)^2$$

Proof. By the definition of R ,

$$\begin{aligned} e_R(a) &= \sum_{\pi} p_R(\pi | a) \pi = \sum_{\pi} \sum_y p(y) p(\pi | a, y) \pi \\ &= \sum_y p(y) E(\pi | a, y) \end{aligned}$$

Because e is pinned down by a (i.e., it is independent of any other variable once we fix a), from now on we can replace any appearance of e (in a term

that holds a fixed) with the notation $e(a)$. Since $\pi = a + \varepsilon$,

$$E(\pi | a, y) = a + E(\varepsilon | a, y)$$

Because $y = \gamma\pi - e + \eta$, we have

$$\gamma\varepsilon + \eta = y - \gamma a + e(a)$$

For given a and y , the R.H.S is a constant, whereas the L.H.S is a sum of two independent variables that are normally distributed with mean zero (and recall that the variance of $\gamma\varepsilon$ is $\gamma^2\sigma_\varepsilon^2$). Therefore, to calculate $E(\varepsilon | a, y)$, we can apply the standard formula for $E(X | X + Y)$ when X and Y are independent normally distributed variables, and obtain

$$E(\varepsilon | a, y) = \frac{\beta}{\gamma}(y - \gamma a + e(a))$$

where

$$\beta = \frac{\gamma^2\sigma_\varepsilon^2}{\gamma^2\sigma_\varepsilon^2 + \sigma_\eta^2} \quad (8)$$

We can now write

$$\begin{aligned} e(a) &= \sum_y p(y) \left[a + \frac{\beta}{\gamma}y - \beta a + \frac{\beta}{\gamma}e(a) \right] \\ &= a(1 - \beta) + \frac{\beta}{\gamma}e(a) + \frac{\beta}{\gamma}\mu_y \end{aligned}$$

Plugging the Phillips curve, we obtain

$$e(a) = (1 - \beta)a + \frac{\beta}{\gamma}e(a) + \frac{\beta}{\gamma}[\gamma\mu_a - E(e(a))]$$

This functional equation defines $e(a)$. Taking expectations, we obtain

$$E(e(a)) = (1 - \beta)\mu_a + \frac{\beta}{\gamma}E(e(a)) + \beta\mu_a - \frac{\beta}{\gamma}E(e(a))$$

such that $E(e(a)) = \mu_a$. This enables us to get the following explicit solution

for $e(a)$:

$$e(a) = \frac{\gamma - \gamma\beta}{\gamma - \beta}a + \frac{\gamma\beta - \beta}{\gamma - \beta}\mu_a$$

Plugging the expression for β , we obtain

$$e(a) = \frac{\sigma_\eta^2}{\gamma(\gamma - 1)\sigma_\varepsilon^2 + \sigma_\eta^2}a + \frac{\gamma(\gamma - 1)\sigma_\varepsilon^2}{\gamma(\gamma - 1)\sigma_\varepsilon^2 + \sigma_\eta^2}\mu_a \quad (9)$$

We have thus pinned down $e_R(a)$ as a function of the strategy that CB chooses, because that strategy induces $\mu_a = \sum_\theta p(\theta)E(a | \theta)$. We can thus restate CB's objective function: choose a strategy (i.e., a stochastic mapping from θ to a) that minimizes

$$\text{Var}(y) + kE(\pi - \theta)^2$$

subject to the constraints

$$\begin{aligned} \pi &= a + \varepsilon \\ y &= \gamma\pi - e(a) + \eta \end{aligned}$$

Plugging (9) into $e(a)$, we can write

$$E(y | a) = (\gamma - \delta)a - (1 - \delta)\mu_a$$

such that

$$\mu_y = (\gamma - 1)\mu_a$$

where

$$\delta = \frac{\sigma_\eta^2}{\gamma(\gamma - 1)\sigma_\varepsilon^2 + \sigma_\eta^2}$$

Because ε and η are independent variables with mean zero, we can ignore them in the calculation of the objective function, which is reduced to

$$(\gamma - \delta)^2 E[(a - \mu_a)^2] + kE[a - \theta]^2 \quad (10)$$

This is exactly the same as in the rational-expectations case, except that

the coefficient $(\gamma - \delta)^2$ replaces $(\gamma - 1)^2$. The policy that minimizes this expression is $a^{**}(\theta)$, as given in the statement of the proposition. Again, the derivation is standard and therefore omitted. ■

This result has a few noteworthy features. First, the expression for $e_R(a)$ given by (9) implies that when $\gamma = 1$, PS has effectively rational expectations - i.e., $e_R(a) = a$ for every a - despite having an incorrect causal model. As a result, the optimal policy under R coincides with the rational-expectations prediction when $\gamma = 1$ - namely, it fully tracks θ .

Deviations from the rational-expectations prediction occur only when $\gamma > 1$. In this case, PS's inflation forecast is a weighted average of a and its ex-ante expected value μ_a . This means that PS's expectations are unbiased on average, and CB would not be able to enhance expected output if it cared about this motive. However, PS's expectations are non-rational, in the sense that they are not fully responsive to fluctuations in CB's actions. The intuition is the same as in the previous sub-section: PS erroneously regards y as an exogenous variable that affects π , and therefore assigns some weight to the ex-ante expected value of y when forming its inflation forecast. Because y is in fact a consequence of a , PS ends up assigning weight to μ_a , thus failing to fully condition on the actual realization of a . The extent of this failure depends on the relative magnitudes of σ_ε^2 and σ_η^2 . As the Phillips relation becomes more reliable (or, equivalently, as the effect of monetary policy on inflation becomes less reliable), the erroneous weight on μ_a increases and PS's deviation from rational expectations is exacerbated.

PS's "expectational rigidity" impels CB toward a more rigid policy relative to the rational-expectations benchmark. This can be immediately seen from the effective objective function (10). Since $\delta \leq 1$ by definition, CB places a larger weight on the consideration of minimizing the variance of a , compared with the rational-expectations benchmark. Excess rigidity of CB's optimal policy increases with $\sigma_\varepsilon^2/\sigma_\eta^2$.

4 General Analysis

Section 3 demonstrated various ways in which CB's optimal policy may depart from the rational-expectations benchmark. In particular, we saw that the optimal policy may coincide with the rational-expectations prediction even when PS's causal model is misspecified. However, the examples involve specific assumptions regarding R and the exogenous components of the true distribution. In this section I look for general conditions on R , for which CB's optimal policy will not depart from the rational-expectations prediction, for *all* possible specifications of the exogenous processes.

Let us first examine the case of $k_v = 0$ - i.e., CB does not care about output variance, and only weighs expected output against the cost of deviating from the inflation target. Normalize $k_e = 1$, and denote $k_t = k$. Plugging Phillips curve (3), the objective function becomes

$$\begin{aligned} & \sum_{\theta} p(\theta) \sum_a p(a | \theta) \cdot [\gamma E(\pi | a) - e_R(a) - kE((\pi - \theta)^2 | \theta, a)] \\ = & \gamma E(\pi) - \sum_a p(a) e_R(a) - kE((\pi - \theta)^2) \end{aligned}$$

When $\gamma = 1$, CB's trade-off is as follows. On one hand, it wishes to deceive PS by causing it to systematically underestimate expected inflation. On the other hand, to the extent that such deception is possible via enhancing fluctuations in π , this has a downside because of the increased deviations from the inflation target.

Recall that when $R = R^*$ - i.e., when PS has rational expectations, the second term in CB's effective objective function coincides with $E(\pi)$, such that CB effectively maximizes

$$(\gamma - 1)E(\pi) - kE((\pi - \theta)^2)$$

The following result provides a sufficient condition for the rational-expectations prediction to persist.

Proposition 5 *Suppose that R includes no v -colliders into the nodes a , π ,*

or into any other node along a directed path that connects a and π . Then,

$$\sum_a p(a)p_R(\pi | a) = p(\pi)$$

for every p . As a result, if $k_v = 0$, then CB's optimal policy is the same as when PS has rational expectations.

Proof. First, it is easy to verify from (5) that if a node is an ancestor of neither a nor π , it can be ignored in the calculation of $p_R(\pi | a)$, hence we can remove it from R without changing our analysis. The proof now proceeds by distinguishing among three cases.

First, suppose there exists no directed path between a and π . Then, by Remark 1, $p_R(\pi | a) = p_R(\pi)$. By assumption, R includes no v -collider into π . Since there are only two nodes apart from a and π , it must be the case that π is ancestral in some DAG in the equivalence class of R . Therefore, $p_R(\pi) = p(\pi)$, and the claim trivially holds.

Second, suppose that $a\tilde{R}\pi$. By assumption, R has no v -collider into a or π . By Proposition 1, this means that $\{a, \pi\}$ is an ancestral clique in some DAG in the equivalence class of R . Therefore, by Remark 2, $p_R(a, \pi) = p(a, \pi)$ and $p_R(a) = p(a)$. Therefore,

$$p_R(\pi | a) = \frac{p_R(a, \pi)}{p_R(a)} = \frac{p(a, \pi)}{p(a)} = p(\pi | a)$$

which again trivially implies the claim.

The third case is the most involved one: assume a and π are not directly linked, but there is a directed path between them. The path consists of one or two nodes apart from a and π . By assumption, there is no v -collider into any node along this path. Therefore, if there are multiple directed paths between a and π , then it must be the case R contains no v -colliders at all, and in this case we can switch to an equivalent DAG that contains a single directed path between a and π . Therefore, we will focus on the latter case only. If the DAG contains a node j that does not lie along the single path between a and π , then either j is an ancestor of neither a nor π (in which case

we noted above that j can be ignored), or jRa and $jR\pi$, or $jR\pi$ and jRa . In all three cases, the relevant part of R contains no v -collider. By Proposition 1, any one of the nodes along the directed path between a and π is ancestral in some DAG in the equivalence class of R . Without loss of generality, then, suppose that the path flows from a into π . Denote the immediate ancestor of π by i . Then, we can write

$$\begin{aligned} \sum_a p(a)p_R(\pi \mid a) &= \sum_a p(a) \sum_{x_i} p_R(x_i \mid a)p(\pi \mid x_i) \\ &= \sum_a p(a) \sum_{x_i} \frac{p_R(x_i)p_R(a \mid x_i)}{p_R(a)} p(\pi \mid x_i) \end{aligned}$$

Because a and i are ancestral in some DAG in the equivalence class of R , Remark 2 implies $p_R(a) = p(a)$ and $p_R(x_i) = p(x_i)$. Therefore,

$$\begin{aligned} \sum_a p(a)p_R(\pi \mid a) &= \sum_a p(a) \sum_{x_i} \frac{p(x_i)p_R(a \mid x_i)}{p(a)} p(\pi \mid x_i) \\ &= \sum_{x_i} p(x_i)p(\pi \mid x_i) \sum_a p_R(a \mid x_i) \\ &= \sum_{x_i} p(x_i)p(\pi \mid x_i) = p(\pi) \end{aligned}$$

which implies the claim. This completes the proof. ■

This result shows that when $k_v = 0$ - i.e., when CB does not care about output variance - the rational-expectations prediction is quite robust to relaxing the assumption that PS has rational expectations. In other words, the feature that CB is unable to systematically fool PS holds for many misspecified PS causal models.

The sufficient condition places a restriction on causal models that postulate a certain variable x_i to be a immediate consequence of two independent, or indirectly linked causes, x_j and x_k . The condition does allow for such configurations, as long as x_i is not a , π or any other variable along a perceived causal chain between a and π . Thus, the following DAGs satisfy the

sufficient condition:

$$\begin{array}{ccc}
 \theta & \rightarrow & a \\
 \theta \rightarrow a \rightarrow y \rightarrow \pi & & \downarrow \quad \downarrow \\
 & & \pi \rightarrow y
 \end{array}$$

The left-hand DAG postulates that CB's action directly determines output, which in turn directly determines inflation. The right-hand DAG postulates that inflation is directly caused by the state of Nature, and that output is jointly determined by CB's action and by inflation.

The sufficient condition is not necessary for PS to hold an unbiased inflation forecast. Let

$$\begin{array}{ccc}
 \theta & \rightarrow & a \\
 R : \downarrow & & \downarrow \\
 \pi & \leftarrow & y
 \end{array}$$

The economic interpretation of R is that the state of Nature is a direct cause of inflation, and that there is another causal chain from CB's policy and inflation, which runs through output. Note that R violates the sufficient condition because it includes a v -collider into π . To see why it leads to unbiased inflation forecasts nonetheless, observe that R is equivalent to a DAG that inverts the causal link between a and θ . Therefore, we can write

$$\sum_a p(a) p_R(\pi | a) = \sum_a p(a) \sum_\theta \sum_y p(\theta | a) p(y | a) p(\pi | \theta, y)$$

Now, according to the true process (or any distribution that is consistent with R^* , for that matter), y is independent of θ conditional on a , hence we

can write $p(\theta | a)p(y | a) = p(\theta, y | a)$. It follows that

$$\begin{aligned}
\sum_a p(a)p_R(\pi | a) &= \sum_a p(a) \sum_\theta \sum_y p(\theta, y | a)p(\pi | \theta, y) \\
&= \sum_a p(a) \sum_\theta \sum_y \frac{p(\theta, y)p(a | \theta, y)}{p(a)} p(\pi | \theta, y) \\
&= \sum_\theta \sum_y p(\theta, y)p(\pi | \theta, y) \sum_a p(a | \theta, y) \\
&= p(\pi)
\end{aligned}$$

This argument relies on the specific structure of R^* , whereas the proof of the Proposition 5 does not rely on any restriction of the true distribution p . I will revisit this observation in Section 5.

Proposition 5 has sharp implications for the case in which θ is *constant*. In this case, the node θ can be omitted from R altogether, such that effectively R is defined over three nodes: a, π, y .

Corollary 3 *Let $k_v = 0$, and suppose that θ is constant. Then, CB's policy is the same as in the rational-expectations benchmark whenever R is not $a \rightarrow \pi \leftarrow y$ or $\pi \rightarrow a \leftarrow y$.*

Proof. The DAGs $a \rightarrow \pi \leftarrow y$ or $\pi \rightarrow a \leftarrow y$ are the only ones that contain v -colliders into a or π . Because there are only three variables, there can be no v -collider into a node along a directed path between a and π . Therefore, any other DAG satisfies the sufficient condition of Proposition 5. ■

In Section 3.2, we saw an example in which CB's optimal strategy departs from the rational-expectations benchmark when the sufficient condition is violated. Another feature of that example was CB employed a random strategy. The following result shows that the two features are logically related.

Proposition 6 *Let $k_v = 0$, and suppose that θ is constant. If CB's optimal strategy departs from the rational-expectations prediction, then it must involve randomization.*

Proof. Suppose CB plays some action a with probability one. Then, because a does not fluctuate, we can omit the node a from R . We only need to consider the subgraph over the nodes π and y that is induced by R . Because this subgraph consists of two nodes, π is an ancestral node across its equivalence class. By Remark 2, $p_R(\pi) = p(\pi)$, which implies the claim. ■

Thus, when θ is constant, departure from the rational-expectations prediction necessarily involves unpredictable CB behavior, because randomization is necessary for creating systematically biased inflation forecasts.

Persistence of the rational-expectations prediction becomes even more extreme when we impose some parametric structure on the exogenous components of p . Recall the linear-normal specification of Section 3.3, according to which $\pi = a + \varepsilon$, where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ and $\eta \sim N(0, \sigma_\eta^2)$. It turns out that under this specification, the rational-expectations prediction for the case of constant θ and $k_v = 0$ persists for *any* causal model that PS might have.

Proposition 7 *Let $\pi = a + \varepsilon$, where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ and $\eta \sim N(0, \sigma_\eta^2)$ are independent. Let $k_v = 0$, and suppose θ is constant. Then, CB's optimal policy is the same as when PS has rational expectations, for any R .*

Proof. By Corollary 3, we only need to consider the case when R is $a \rightarrow \pi \leftarrow y$ or $\pi \rightarrow a \leftarrow y$. Suppose that $R : a \rightarrow \pi \leftarrow y$. In the proof of Proposition 4, we established that in this case, $e_R(a)$ is given by (9). Since $e_R(a)$ is a weighted average of a and μ_a , it immediately follows that $\sum_a p(a)e_R(a) = \mu_a = \mu_\pi$ (the latter identity follows from the assumption that $\pi = a + \varepsilon$ and ε is independent with mean zero).

Now suppose that $R : \pi \rightarrow a \leftarrow y$. Then,

$$\begin{aligned} e_R(a) &= \sum_{\pi} p_R(\pi | a)\pi = \sum_{\pi} \sum_y p(y | a)p(\pi | a, y)\pi \\ &= \sum_y p(y | a)E(\pi | a, y) \end{aligned}$$

Calculating $E(\pi | a, y)$ proceeds exactly as in the proof of Proposition 4.

Following essentially the same ensuing steps in the proof, we obtain

$$e_R(a) = a(1 - \beta) + \frac{\beta}{\gamma}e_R(a) + \frac{\beta}{\gamma}E(y | a)$$

where β is given by (8). Plugging the Phillips Curve, $E(y | a) = \gamma a - e_R(a)$, we obtain $e_R(a) = a = E(\pi | a)$. That is, PS has rational expectations. It follows that the rational-expectations prediction persists for any R . ■

This is perhaps the most extreme robustness result in this paper. When θ is constant and BC is only guided by the output-enhancing and inflation-targeting motives, its optimal policy is the same for all possible PS causal models, as long as we impose the conventional linear-normal parametric restriction on $p(\pi | a)$ and the Phillips Curve.

Let us now allow k_v to take positive values, such that CB is guided by all three motives. In this case, the sufficient condition for the persistence of the rational-expectations prediction is more stringent.

Proposition 8 *Let $k_v \geq 0$. Suppose that $a\tilde{R}\pi$ and R has no v -collider into a or π . Then, CB's optimal policy is the same as when PS has rational expectations, for any specification of the exogenous components of p .*

Proof. In the proof of Proposition 5, we saw that if $a\tilde{R}\pi$ has no v -collider into a or π , then $p_R(\pi | a) \equiv p(\pi | a)$. In this case, CB's effective objective function is the same as in the rational-expectations benchmark. ■

Thus, when CB also cares about output variance, persistence of the rational-expectations prediction does not rely merely on CB's inability to systematically fool PS; it also requires PS to correctly perceive the correlation between inflation and CB's action. This places additional restrictions on PS's departure from rational expectations. The DAG $\theta \rightarrow y \leftarrow a \leftarrow \pi$, for example, satisfies the sufficient condition, although it deviates from R^* in many ways.

To illustrate a DAG that violates the sufficient condition of Proposition 8, while satisfying the sufficient condition of Proposition 5, consider $R : \theta \rightarrow a \rightarrow y \leftarrow \pi$. Note that the nodes a and π are not directly linked. As in

Section 3.3, let $k_e = 0$, $k_v = 1$, $k_t = k$, and assume $\pi = a + \varepsilon$, where ε is an independent zero-mean random variable. In this case, $e_R(a)$ is equal to μ_π for every a . Suppose that $\gamma = 1$. Then, CB's objective function is reduced to minimizing

$$E[(a - \mu_a]^2 + kE[a - \theta]^2$$

such that the optimal solution is

$$a(\theta) = \frac{k}{1+k}\theta + \frac{1}{1+k}\mu_\theta$$

whereas the rational-expectations solution is $a^*(\theta) = \theta$. That is, PS's misspecified causal model leads CB to adopt a relatively rigid policy.

5 Variations and Extensions

In this section I briefly discuss a few variations and extensions of the model.

5.1 The No-Commitment Regime

Throughout the paper, I assumed that CB commits ex-ante to its policy. Of course, the original Kydland-Prescott and Barro-Gordon models were developed to highlight the role of commitment when PS has rational expectations. However, note that in this paper, I assumed that PS *observes* CB's actions. If PS had rational expectations, there would be no role for ex-ante commitment, because CB would never be tempted to deviate from the ex-ante optimal action: PS would be able to monitor any deviation from the pre-committed action and adapt its rational inflation forecast accordingly.

In contrast, when PS has a misspecified causal model, a commitment problem does arise even when PS perfectly monitors CB's actions. Suppose that $R : \theta \rightarrow a \rightarrow y \leftarrow \pi$, and let $k_v = 0$. We saw in Section 4 that in this case, $p_R(\pi | a) = p(\pi)$ for every a - i.e., PS's inflation forecast is unbiased on average, yet at the same time it is entirely unresponsive to the realization of a . In other words, PS behaves as if it has rational expectations but cannot monitor CB's action - exactly as in the original Kydland-Prescott and Barro-

Gordon models! Therefore, the analysis of CB's ex-ante and time-consistent policies in this example is reduced to the standard treatment.

5.2 Alternative True Processes

The true process I assumed in this paper is consistent with the DAG R^* described in Section 2. According to this causal model, inflation is determined by CB's action only, and output is a function of inflation and inflationary expectations. This formulation, due to Sargent (2001), departs from the original Barro-Gordon model, where output and inflation are determined jointly as a function of CB's action and inflationary expectations.

The main sufficiency results of Section 4 (with the exception of Proposition 7) are absolutely robust to changes in the specification of the true process that governs output and inflation. The reason is that these results state sufficient conditions for persistence of the rational-expectations prediction for *any* specification of the exogenous components of p . Nowhere in the proofs of these results did we rely on any structure that these components might have. In particular, for these results we could make any assumption about how π and y are determined as a function of a . When we turn to calculations of specific examples, the Sargent formulation is at times easier to work with.

5.3 Additional Variables

Throughout the paper, I assumed that the true model - and PS's causal model - involve no more than five variables (θ, a, e, π, y) . It would be natural to extend the model by adding variables. For example, CB may condition its action on a collection of observable variables. And the mapping from CB's action to inflation and output may involve additional variables. Incorporating additional variables will be crucial as we consider dynamic extensions of the present model.

The sufficient conditions of Section 4 can be adapted to environments with additional variables. In particular, the following is an extension of the sufficient condition in Proposition 5: R should include no v -collider into a

node along *any* directed path into a or π (rather than just directed paths between these two nodes). Any DAG R that satisfies this condition will generate inflation forecasts that are unbiased on average.

5.4 Heterogenous Causal Models

Throughout the paper, I assumed that CB knows PS's DAG R . The model can be extended to allow for uncertainty in this regard. Suppose that CB assigns probability $\lambda(R)$ to every DAG R . For every realized R , $e_R(a)$ is defined as before, such that the average inflation forecast conditional on a is

$$e(a) = \sum_R \lambda(R) \sum_{\pi} p_R(\pi | a) \pi$$

The linearity of the Phillips Curve implies that expected output conditional on a is $\gamma E(\pi | a) - e(a)$. As a result, extending the analysis to the case of an uncertain DAG is straightforward. In particular, the general results in Section 4 can be slightly restated, such that *every* DAG in the support of λ is required to satisfy the sufficient conditions.

6 Conclusion

If a central bank could cause the private sector to systematically underestimate inflation, it might be able to use monetary policy to exploit the empirical output-inflation relation, in order to obtain systematic enhancement of real output. The rational-expectations revolution in 1970s macroeconomics discredited this idea, by arguing that if the private sector has rational expectations, it cannot be systematically fooled. In this paper, I revisited this theme, and showed that results in a similar vein are obtained under the substantially weaker assumption that the private sector forms inflationary expectations by fitting a misspecified causal model to long-run data. However, this does not mean that non-rational expectations are irrelevant for monetary policy, because they may impel the central bank to adopt a rigid policy, which does not track changes in the ideal inflation target as much as

it would under private-sector rational expectations.

References

- [1] Athey, S., A. Atkeson and P. Kehoe (2005), “The Optimal Degree of Discretion in Monetary Policy,” *Econometrica* 73, 1431-1475.
- [2] Barro, R. and D. Gordon (1983), “Rules, Discretion and Reputation in a Model of Monetary Policy,” *Journal of Monetary Economics* 12, 101-121.
- [3] Cho, I., N. Williams and T. Sargent (2002), “Escaping Nash Inflation,” *Review of Economic Studies*, 69, 1-40.
- [4] Cowell, R., P. Dawid, S. Lauritzen and D. Spiegelhalter (1999), *Probabilistic Networks and Expert Systems*, Springer, London.
- [5] Esponda, I. and D. Pouzo (2015), “Berk-Nash Equilibrium: A Framework for Modeling Agents with Misspecified Models,” mimeo.
- [6] Evans, G. and S. Honkapohja (2001), *Learning and Expectations in Macroeconomics*, Princeton University Press.
- [7] Garcia-Schmidt, M. and M. Woodford (2015), “Are Low Interest Rates Deflationary? A Paradox of Perfect-Foresight Analysis,” NBER Working Paper no. w21614.
- [8] Giacomini, R., V. Skreta and J. Turen (2015), “Models, Inattention and Expectation Updates,” CEPR Discussion Paper no. 11004.
- [9] Hoover, K. (2001), *Causality in macroeconomics*, Cambridge University Press.
- [10] Klamler, A. (1984), *The New Classical Macroeconomics: Conversations with the New Classical Economists and their Opponents*. Wheatsheaf Books.

- [11] Kydland, F. and E. Prescott (1977), “Rules rather than Discretion: The Inconsistency of Optimal Plans,” *Journal of Political Economy* 85, 473-491.
- [12] Lucas, R. (1972), “Expectations and the Neutrality of Money,” *Journal of Economic Theory* 4, 103-124.
- [13] Pearl, J. (2009), *Causality: Models, Reasoning and Inference*, Cambridge University Press, Cambridge.
- [14] Piccione, M. and A. Rubinstein (2003), “Modeling the Economic Interaction of Agents with Diverse Abilities to Recognize Equilibrium Patterns,” *Journal of the European Economic Association* 1, 212-223.
- [15] Sargent, T. (2001), *The conquest of American inflation*, Princeton University Press.
- [16] Sargent, T. and N. Wallace (1975), “‘Rational’ Expectations, the Optimal Monetary Instrument, and the Optimal Money Supply Rule,” *Journal of Political Economy* 83, 241-254.
- [17] Spiegel, R. (2015a), “Bayesian Networks and Missing-Data Imputation,” *Quarterly Journal of Economics*, forthcoming.
- [18] Spiegel, R. (2015b), “On the ‘Limited-Feedback’ Foundation of Boundedly Rational Expectations,” mimeo.
- [19] Verma, T. and J. Pearl (1991), “Equivalence and Synthesis of Causal Models,” *Uncertainty in Artificial Intelligence*, 6, 255-268.
- [20] Woodford, M. (2013), “Macroeconomic Analysis without the Rational Expectations Hypothesis,” *Annual Review of Economics*, forthcoming.