Macroeconomic Analysis Without the Rational Expectations Hypothesis

Michael Woodford

Department of Economics, Columbia University, New York, NY 10027; email: michael.woodford@columbia.edu

Annu. Rev. Econ. 2013. 5:303–46

First published online as a Review in Advance on April 29, 2013

The *Annual Review of Economics* is online at economics.annualreviews.org

This article's doi:
10.1146/annurev-economics-080511-110857

JEL codes: D84, E52

Keywords

temporary equilibrium, learning dynamics, Ricardian equivalence, deflation trap, Taylor rule

Abstract

The article presents a temporary equilibrium framework for macroeconomic analysis that allows for a wide range of possible specifications of expectations but reduces to a standard new Keynesian model in the limiting case of rational expectations. This common framework is then used to contrast the assumptions and implications of several different ways of relaxing the assumption of rational expectations. As an illustration of the method, the implications of alternative assumptions for the selection of a monetary policy rule are discussed. Other issues treated include the conditions required for Ricardian equivalence and for existence of a deflation trap.

1. INTRODUCTION

A crucial methodological question in macroeconomic analysis is the way in which decision makers' expectations about future conditions should be modeled. To the extent that behavior is modeled as goal directed, it will depend (except in the most trivial cases) on expectations, and analyses of the effects of alternative governmental policies need to consider how expectations are endogenously influenced by one policy or another. Finding tractable ways to address this issue has been a key challenge for the extension of optimization-based economic analysis to the kinds of dynamic settings required for most questions of interest in macroeconomics.

The dominant approach for the past several decades of course has made use of the hypothesis of model-consistent or rational expectations (RE): the assumption that people have probability beliefs that coincide with the probabilities predicted by one's model. The RE benchmark is a natural one to consider, and its use has allowed a tremendous increase in the sophistication of the analysis of dynamics in the theoretical literature in macroeconomics. Nonetheless, the assumption is a strong one, and one may wonder if it should be relaxed, especially when considering relatively short-run responses to disturbances, or the consequences of newly adopted policies that have not been followed in the past—both of which are precisely the types of situations that macroeconomic analysis frequently seeks to address.

Although the assumption that an economy's dynamics must necessarily correspond to a rational expectations equilibrium (REE) may seem unjustifiably strong—and, under some circumstances, is a heroic assumption indeed—it does not follow that we should then be equally willing to entertain all possible assumptions about the expectations of economic agents. It makes sense to assume that expectations should not be completely arbitrary and have no relation to the kind of world in which the agents live; indeed, it is appealing to assume that people's beliefs should be rational, in the ordinary-language sense, although there is a large step from this to the RE hypothesis.[1] We should therefore like to replace the RE hypothesis by some weaker restriction that nonetheless implies a substantial degree of conformity between people's beliefs and reality—that implies, at the least, that people do not make obvious mistakes.

The literature has explored two broad approaches to the formulation of a criterion for reasonableness of beliefs that is weaker than the RE hypothesis. One is to assume that people should correctly understand the economic model and be able to form correct inferences from it about possible future outcomes. The other is to assume that the probabilities that people assign to possible future outcomes should not be too different from the probabilities with which different outcomes actually occur, given that experience should allow some acquaintance with these probabilities, regardless of whether people understand how these outcomes are generated. The former approach supposes that beliefs should be refined through a process of reflection, independent of experience and not necessarily occurring in real time, that Guesnerie (1992) calls "eduction," whereas the latter supposes that beliefs should be refined over time through a process of induction from observed outcomes. Section 3 discusses the first approach, and examples of the second approach are taken up in Sections 4 and 5.

Within the category of inductive approaches, one may distinguish two important subcategories. A first class of approaches specifies the beliefs that should be regarded as reasonable by predicting the patterns that people should be able to recognize in the data on the basis of the rationality of the procedure used to look for such patterns. A different class of approaches specifies a degree of correspondence between subjective and model-implied probabilities that should be

[1]This is stressed, for example, by Kurz (2012, p. 1).

expected, without explicitly modeling the process of inference through which such beliefs are formed. The first class of approaches (models of econometric learning) is treated in Section 4, and the second class (theories of partially or approximately correct beliefs) is taken up in Section 5.

The different possible approaches to the specification of expectations are compared by illustrating the application of each in the context of the same general framework for macroeconomic analysis, introduced in Section 2. In each case, it is shown that one can demand that the specification of beliefs satisfy quite stringent rationality requirements without, in general, being able to conclude that the predictions of the REE analysis must obtain. I consider in particular the consequences of alternative specifications of expectations for several familiar issues: the conditions under which an interest-rate rule for monetary policy should be able to maintain a stable inflation rate, the nature of the trade-off between inflation stabilization and output stabilization, and the effects of the government budget on aggregate demand.

2. TEMPORARY EQUILIBRIUM IN A NEW KEYNESIAN MODEL

I begin by introducing a framework for analysis of the determinants of aggregate output and inflation, in which subjective expectations can be specified arbitrarily, to clarify the role of alternative specifications of expectations. It allows the effects of both monetary and fiscal policies to be considered, along with a variety of types of exogenous disturbances to economic fundamentals and possible shifts in expectations. The model is one in which households and firms solve infinite-horizon optimization problems, as in the DSGE (dynamic stochastic general equilibrium) models commonly used for quantitative policy analysis; in fact, under the RE assumption, the model presented here corresponds to a textbook new Keynesian (NK) model of the kind analyzed in Clarida et al. (1999), Woodford (2003, chapter 4), Gali (2008, chapter 3), and Walsh (2010, chapter 8). Essentially, the goal of this section is to show how temporary equilibrium (TE) analysis of the kind introduced by Hicks (1939) and Lindahl (1939) and further developed by Grandmont (1977, 1988)—in which a general competitive equilibrium is defined at each point in time, on the basis of the (independently specified) expectations that decision makers happen to entertain at that time—can be extended to a setting with monopolistic competition, sticky prices, and infinite-horizon planning for closer comparison with the conclusions of conventional macroeconomic analysis. Some of the best-known conclusions from RE analysis of the model are also recalled as a basis for comparison with the conclusions from alternative specifications of expectations in the later sections.

2.1. Expectations and Aggregate Demand

The economy is made up of a large number of identical households, and a large number of firms, each of which is the monopoly producer of a particular differentiated good, with each household owning an equal share of each firm.[2] At any point in time, a household has an infinite-horizon consumption (and wealth-accumulation) plan from that date forward, which maximizes expected discounted utility under certain subjective expectations regarding the future evolution of income and the rate of return on saving, and the household's expenditure at that date is assumed to be the one called for by the plan believed to be optimal at that time. (The household may or may not continue to execute the same plan as time passes, depending on what is assumed about how expectations change.) Although I wish for now to leave the subjective expectations unspecified, the

[2]These and other aspects of the model structure are explained in more detail in Woodford (2003). Here I focus only on the points at which an alternative model of expectations requires a generalization of the standard exposition.

expectations held at any date represent a well-behaved probability measure over possible future evolutions of the state variables. Because I do not assume that subjective expectations are necessarily model consistent, they are not necessarily the same across households, nor do I necessarily assume that a household's later expectations are those that it previously expected to hold.

To simplify, I assume that the only traded asset is riskless nominal one-period government debt.[3] I further assume that households have no choice but to supply the hours of work that are demanded by firms, at a wage that is fixed by a union that bargains on behalf of the households. A household then has a single decision each period, which is the amount to spend on consumption. Because its nonfinancial income (the sum of its wage income and its share of the profits of the firms) is outside its control, in order to analyze optimal expenditure, we need to specify only the household's expectations regarding the time path of its total nonfinancial income. If an equal amount of work is demanded from each household at the wage fixed by the union, and households similarly each receive an equal share of the profits of all the firms,[4] then each household's nonfinancial income will be the same each period and will be equal to its share of the total value of output in that period; hence we can equivalently specify nonfinancial income expectations as expectations regarding the path of output.

A household's perceived intertemporal budget constraint then depends only on its expectations about the path of aggregate output, the path of aggregate tax collections (also assumed to be levied equally on each household), and the real return on the one-period debt. The consumption planning problem for an individual household at a given point in time is then the familiar one faced by a household with RE and an exogenously given income process (discussed, e.g., in Deaton 1992), and the solution is correspondingly the same, except with subjective probabilities substituted for objective ones.[5]

I log-linearize this and other structural relations of the model around a deterministic steady state, in which (*a*) all exogenous state variables are forever constant, (*b*) monetary and fiscal policy are specified to maintain a constant zero rate of inflation and some constant positive level of public debt, and (*c*) all subjective expectations are correct (i.e., households and firms have perfect foresight). The log-linearized relations accordingly represent an approximation to the exact model, applying in the case in which exogenous disturbances are sufficiently small, monetary and fiscal policies are sufficiently close to being consistent with this steady state, and expectational errors are sufficiently small.

A log-linear approximation to the consumption function takes the form

$$\begin{aligned} c_t^i = (1-\beta)b_t^i + \sum_{T=t}^{\infty}\beta^{T-t}\,\hat{E}_t^i\Big\{&(1-\beta)(Y_T-\tau_T) - \beta\sigma(i_T-\pi_{T+1}) \\ &+(1-\beta)s_b(\beta i_T-\pi_T) - \beta\left(\bar{c}_{T+1}-\bar{c}_T\right)\Big\}. \end{aligned} \tag{1}$$

Here c_t^i is total real expenditure by household i (on all the differentiated goods) in period t, measured as a deviation from the steady-state level of consumption and expressed as a fraction of steady-state

[3]In many RE analyses, with a representative household and fiscal policy assumed to be Ricardian (in the sense defined in Woodford 2001), the model dynamics are unaffected by allowing additional financial markets or even issuance of other types of government debt. The restriction to one-period debt is no longer innocuous, however, once one allows for more general hypotheses regarding expectations, as shown, for example, by Eusepi & Preston (2012b). Nonetheless, I consider only the most simple case here.

[4]Equity ownership shares are assumed for simplicity to be nontradeable.

[5]Note that the log-linear theory of aggregate demand derived here is the same as in a model in which households are assumed simply to have an exogenous endowment of the consumption good, such as that of Guesnerie (2009).

output; b_t^i is the value of maturing bonds carried into period t by household i, deflated by the period $t-1$ price level (rather than the period t price level) so that this wealth measure is predetermined at date $t-1$;[6] Y_t is the deviation of aggregate output from its steady-state value, as a fraction of that steady-state value; τ_t is net tax collections, also measured as a deviation from the steady-state level of tax revenues and expressed as a fraction of steady-state output; i_t is the riskless one-period nominal interest rate on debt issued in period t; π_t is the inflation rate between periods $t-1$ and t; and $\bar{c}_t$ is an exogenous shock to the utility from consumption in period t. In addition, $0<\beta<1$ is the utility discount factor; $\sigma>0$ measures the intertemporal elasticity of substitution; and $s_b>0$ is the steady-state level of government debt, expressed as a fraction of the steady-state output.

The notation $\hat{E}_t^i\{\cdot\}$ indicates the expected value of the terms in the curly brackets, under the subjective expectations of household i in period t. The b_t^i and $Y_T-\tau_T$ terms then represent (a subjective version of) the usual permanent-income hypothesis;[7] the $\sigma(i_T-\pi_{T+1})$ terms indicate how expectations of a real interest rate different from the rate of time preference shift the desired time path of spending; the $s_b(\beta i_T-\pi_T)$ terms indicate the income effects of variations in nominal interest rates and inflation; and the $\bar{c}_T$ terms indicate the effects of preference shocks on the desired time path of spending.

Equation 1 involves subjective expectations of a number of variables at many future horizons, but under our linear approximation we can write desired expenditure as a function of the household's forecast of a single variable,

$$c_t^i=(1-\beta)b_t^i+(1-\beta)(Y_t-\tau_t)-\beta\left[\sigma-(1-\beta)s_b\right]i_t-(1-\beta)s_b\pi_t+\beta\bar{c}_t+\beta\hat{E}_t^i v_{t+1}^i, \qquad (2)$$

where the composite variable v_t^i is defined as

$$v_t^i\equiv\sum_{T=t}^{\infty}\beta^{T-t}\hat{E}_t^i\Big\{(1-\beta)(Y_T-\tau_T)-\left[\sigma-(1-\beta)s_b\right](\beta i_T-\pi_T)-(1-\beta)\bar{c}_T\Big\}. \qquad (3)$$

The advantage of this notation is that we need only to specify how people forecast a single variable each period; note, however, that the variable that people must forecast is a subjective state variable that will depend on their own future forecasts.

Aggregate demand is then given by $Y_t=\int c_t^i di+G_t,$ where G_t is the departure of government purchases of the composite good from their steady-state levels, also measured as a fraction of steady-state output. Substituting Equation 2 for c_t^i in this expression, we obtain

$$Y_t=g_t+(1-\beta)b_t+v_t-\sigma\pi_t, \qquad (4)$$

where $g_t\equiv G_t+\bar{c}_t$ is a composite exogenous disturbance to autonomous expenditure,[8] $b_t\equiv\int b_t^i di$ is the aggregate supply of public debt, and $v_t\equiv\int v_t^i di$ is the average value of the expectational variable defined in Equation 3. We thus obtain an equation for aggregate demand

[6]The reduction in the real value of this wealth due to inflation between periods $t-1$ and t is then reflected in the $-s_b\pi_t$ term inside the curly brackets.

[7]Readers are referred to Deaton (1992), for example, for an exposition of the standard theory, under the assumptions of rational expectations, no preference shocks, and a constant real interest rate.

[8]To simplify, I treat government purchases as an exogenous disturbance rather than as a possibly endogenous policy choice.

that separates out the effects of the exogenous disturbances g_t, the wealth effect of public debt, and the average state of expectation captured by v_t.

The government's flow budget constraint implies a law of motion

$$b_{t+1} = \beta^{-1}[b_t - s_b\pi_t - s_t] + s_b i_t \tag{5}$$

for the public debt, where $s_t \equiv \tau_t - G_t$ is the real primary budget surplus in period t, measured as a deviation from its steady-state level and expressed as a fraction of steady-state output. The aggregate demand block of our model then consists of Equations 4 and 5, together with monetary and fiscal policy equations that specify the evolution of i_t and s_t, respectively (generally as a function of other endogenous variables), and a specification of the evolution of the forecasts $\{\hat{E}^i_t v^i_{t+1}\}$ (which determine the expectational variable v_t). We then have a system of four equations per period (plus the specification of expectations) to determine the paths $\{Y_t, b_t, i_t, s_t\}$ given a path for the price level, or the paths $\{\pi_t, b_t, i_t, s_t\}$ given a supply-determined path for output, along with the composite disturbance g_t and shocks to policy and expectations.

The definition in Equation 3 has a recursive form so that the only subjective expectation involved in v^i_t is i's forecast of the corresponding variable one period in the future. Specifically, Equation 3 implies that

$$v^i_t = (1-\beta)v_t + \beta(1-\beta)(b_{t+1} - b_t) - \beta\sigma(i_t - \pi_t) + \beta\hat{E}^i_t v^i_{t+1}, \tag{6}$$

where in addition to substituting the forecast of v^i_{t+1} for the expectational terms, I have used Equation 4 to substitute out Y_t and Equation 5 to substitute out τ_t. Hence to specify the aspects of subjective expectations that are relevant for aggregate demand determination, it suffices that we specify an evolution for the $\{v^i_t\}$ and subjective one-period-ahead forecasts of those variables that are consistent with Equation 6. This result is used below to characterize the possible TE dynamics under various more specific assumptions about expectations.

2.2. Ricardian Expectations

Equation 5 implies that nonexplosive dynamics for the real public debt require that

$$b_t = \sum_{T=t}^{\infty} \beta^{T-t}\left[s_T - s_b(\beta i_T - \pi_T)\right].$$

I say that households have Ricardian expectations if they expect that the path of primary surpluses will necessarily satisfy this condition so that

$$\hat{E}^i_t \sum_{T=t}^{\infty} \beta^{T-t}\left[s_T - s_b(\beta i_T - \pi_T)\right] = b_t \tag{7}$$

at all times. It is not obvious that expectations must have this property, even under the RE hypothesis;[9] it is even less obvious that reasonable expectations must have this property once one

[9]It is possible for people to believe in a fiscal rule without this property, and yet for an REE to exist, as shown, for example, in Woodford (2001). (In equilibrium, debt does not explode, and expectations are correct, but people believe that debt would explode in the case of certain paths for endogenous variables that do not occur in equilibrium.) Some have disputed whether such a specification of out-of-equilibrium beliefs should be considered to be consistent with the RE hypothesis (see Bassetto 2002 for a careful discussion).

dispenses with the strict RE assumption. Nonetheless, this property is frequently assumed (at least tacitly) in non-RE analyses, and I mainly assume it in the discussion below to simplify the analysis.[10]

Under this assumption, there is no longer a wealth effect of variation in the public debt; the b_t term in Equation 4 is exactly canceled by the effects of the change in the expected path of primary surpluses on the v_t term. In fact, one can write Equation 4 more simply as

$$Y_t = g_t + \overline{v}_t - \sigma\pi_t, \tag{8}$$

where $\overline{v}_t \equiv v_t + (1-\beta)b_t$ is the average value of a subjective variable $\overline{v}_t^i$ which (under the hypothesis of Ricardian expectations) can be defined simply as

$$\overline{v}_t^i \equiv \sum_{T=t}^{\infty} \beta^{T-t}\hat{E}_t^i\{(1-\beta)(Y_T - g_T) - \sigma(\beta i_T - \pi_T)\}. \tag{9}$$

In this case, aggregate demand determination is completely independent of the paths of both public debt and tax collections—so that Ricardian equivalence obtains—as long as these fiscal variables have no direct effect on people's expectations regarding the evolution of the variables $\{Y_t, \pi_t, i_t, g_t\}$ or, more simply, as long they have no direct effect on forecasts of the path of the variables $\{v_t^i\}$.[11]

Under the assumption of Ricardian expectations, it is convenient to specify expectations by describing the evolution of the variables $\{\overline{v}_t^i\}$; these must be consistent with a relation of the form

$$\overline{v}_t^i = (1-\beta)\overline{v}_t - \beta\sigma(i_t - \pi_t) + \beta\hat{E}_t^i\overline{v}_{t+1}^i, \tag{10}$$

analogous to Equation 6. The complete aggregate demand block of the model then consists of Equation 8, a monetary policy equation, and a specification of the evolution of the expectational variable $\{\overline{v}_t^i\}$ consistent with Equation 10. This system provides two equations per period to determine the paths of $\{Y_t, i_t\}$ given the evolution of the price level, the exogenous disturbances $\{g_t\}$, and shocks to policy and expectations.

2.3. Expectations and Aggregate Supply

Each of the monopolistically competitive firms sets the price for the particular good that it alone produces. Prices are assumed to remain fixed for a random interval of time: Each period, fraction $0 < \alpha < 1$ of all goods prices remain the same (in monetary terms) as in the previous period, while the other prices are reconsidered, and the probability that any given price is reconsidered in any period is assumed to be independent of both the current price and the length of time that the price has remained fixed. It then follows that (again in a log-linear approximation) the rate of inflation π_t between periods $t-1$ and t will be given by

[10]In many NK models with adaptive learning, TE relations are derived by simply inserting subjective expectations into equations that hold (in terms of model-consistent conditional expectations) in the RE version of the model; that government liabilities are not perceived to be net wealth in the RE analysis then leads to an assumption that they are not in the learning analysis, without any discussion of the assumption. Evans & Honkapohja (2010) instead make the assumption of Ricardian expectations explicit. Eusepi & Preston (2012a,b) and Benhabib et al. (2012) provide examples of analyses in which Ricardian expectations are not assumed.

[11]Readers are referred to Evans et al. (2012) for a more detailed discussion of the conditions under which Ricardian equivalence obtains even without the RE hypothesis.

$$\pi_t = (1-\alpha)p_t^*, \tag{11}$$

where for each firm j, p_t^{*j} is the amount by which the firm would choose to set the log price of its good higher than p_{t-1}, the log of the general price index in period $t-1$, were it to be one of the firms that reconsiders its price in period t, and $p_t^* \equiv \int p_t^{*j} dj$ is the average value of this quantity across all firms.

Each firm that reconsiders its price in a given period chooses the new price that it believes will maximize the present value of its profits from that period onward. In a Dixit-Stiglitz model of monopolistic competition, with a single economy-wide labor market, profits in any period are the same function of a firm's own price and of aggregate market conditions for each firm. The single-period profit-maximizing log price p_t^{opt} is then the same for each firm, and a log-linear approximation to the first-order condition for optimal price-setting takes the form

$$p_t^{*j} = (1-\alpha\beta)\sum_{T=t}^{\infty}(\alpha\beta)^{T-t}\,\widehat{E}_t^j\, p_T^{opt} - p_{t-1}, \tag{12}$$

where β is again the utility discount factor (also the rate at which real profits are discounted in the steady state), and $\widehat{E}_t^j[\cdot]$ indicates a conditional expectation with respect to the subjective beliefs of firm j at date t. The recursive form of Equation 12 implies that internally consistent expectations on the part of any firm must satisfy

$$p_t^{*j} = \left(1-\alpha\beta\right)\left(p_t^{opt} - p_{t-1}\right) + \alpha\beta\left(\widehat{E}_t^j p_{t+1}^{*j} + \pi_t\right). \tag{13}$$

Averaging this expression over firms j and using the resulting equation to substitute for p_t^* in Equation 11, we see that inflation determination depends only on the average of firms' subjective forecasts of a single variable, namely each firm's own value for the expectational variable p_T^{*j} one period in the future.

Suppose furthermore that the union sets a wage each period with the property that at that wage, a marginal increase in labor demand would neither increase nor decrease average perceived utility across households, if for each household the marginal utility of the additional wage income is weighed against the marginal disutility of the additional work. This implies that (in a log-linear approximation)

$$w_t = \nu_t - \lambda_t,$$

where w_t is the log real wage, ν_t is the log of the (common) marginal disutility of labor, and λ_t is the average across households of λ_t^i, a household's subjective assessment of its marginal utility of additional real income. Because optimizing consumption demand implies that

$$\lambda_t^i = -\sigma^{-1}\left(c_t^i - \bar{c}_t\right),$$

we obtain

$$w_t = \nu_t + \sigma^{-1}\left(c_t - \bar{c}_t\right) = \nu_t + \sigma^{-1}(Y_t - g_t),$$

just as in a representative-household model with RE and a competitive economy-wide labor market. In a model in which labor is the only variable factor of production, both ν_t and the marginal product of labor can be expressed as functions of hours worked and hence as functions of output Y_t and the exogenous level of productivity in period t. We then obtain a relation of the form

$$p_t^{opt} = p_t + \xi\left(Y_t - Y_t^n\right) + \mu_t, \tag{14}$$

where the natural rate of output Y_t^n is a composite exogenous disturbance, involving variations in g_t, productivity, and shocks to the disutility of labor, and μ_t measures exogenous variations in the desired markup of prices over the marginal cost.[12]

Substituting Equation 11 for π_t and Equation 14 for p_t^{opt} in Equation 13, we obtain

$$p_t^{*j} = (1-\alpha)p_t^* + (1-\alpha\beta)\left[\xi y_t + \mu_t\right] + \alpha\beta\widehat{E}_t^j p_{t+1}^{*j}, \tag{15}$$

where y_t indicates the output gap, defined as $Y_t - Y_t^n$. For beliefs to be internally consistent, the evolution of the expectational variables $\{p_t^{*j}\}$ and subjective one-period-ahead forecasts of those variables must satisfy Equation 15 at all times. Given such beliefs, the inflation rate will be determined by Equation 11. Thus Equations 11 and 15 constitute the aggregate-supply block of the model, which determines the evolution of the general price level, given the evolution of output, exogenous disturbances, and expectations.

2.4. The Complete Model

In the complete TE system, the expectations that must be specified are paths for $\{\overline{v}_t^i\}$ for all households and $\{p_t^{*j}\}$ for all firms, consistent with the relations in Equations 10 and 15, respectively. Given these expectations, paths for the exogenous disturbances $\{g_t, Y_t^n, \mu_t\}$, and a monetary policy rule for the evolution of $\{i_t\}$, the evolutions of aggregate output and inflation are then given by Equations 8 and 11.

Substituting for the variables v_t^i and p_t^{*j} in terms of current observables and forecasts of future conditions, Equation 8 can be written as

$$Y_t = g_t - \sigma i_t + \int \widehat{E}_t^i \overline{v}_{t+1}^i di, \tag{16}$$

and Equation 11 can correspondingly be written as

$$\pi_t = \kappa y_t + u_t + (1-\alpha)\beta \int \widehat{E}_t^j p_{t+1}^{*j} dj, \tag{17}$$

where

$$\kappa \equiv \frac{(1-\alpha)(1-\alpha\beta)}{\alpha}\xi > 0, \quad u_t \equiv \frac{(1-\alpha)(1-\alpha\beta)}{\alpha}\mu_t.$$

We thus obtain an IS equation and AS equation to describe short-run output and inflation determination, given monetary policy, exogenous disturbances, and expectations.

If expectations are assumed not to change in response to policy changes or other shocks, the model makes predictions similar to those of a standard undergraduate textbook exposition. For example, an increase in the central bank's interest-rate target i_t should reduce output and inflation.

[12]The separation of the effects of exogenous disturbances into Y_t^n and μ_t components in Equation 14 is useful only in the case in which stabilization of the output gap is a goal of policy.

An increase in government purchases (which increases g_t, and also Y_t^n by a smaller amount) should increase both output and inflation. Moreover, a cost-push shock $u_t > 0$ should increase inflation but should have no effect on output if the central bank's interest-rate target is unchanged; if i_t is instead raised in response to the increase in inflation, output should fall, and the inflation increase will be more modest.

However, the model also indicates the effects on output and inflation of changes in average expectations. For example, an increase in the average forecast by firms of the log price that they would wish to choose if reviewing their prices one period in the future, relative to the current average price, should raise current inflation for any current level of output, just as in the case of an exogenous cost-push shock. Additionally, there is no general reason to suppose that expectations should be unaffected by shocks of the kind considered in the previous paragraph; hence a complete analysis even of short-run equilibrium requires a specification of how expectations are determined.

If expectations are Ricardian and monetary policy is unaffected by fiscal variables, the model implies that neither the size of the public debt nor the government budget matters for output and inflation determination. If instead expectations are not assumed to be Ricardian, it is more convenient to write the model in terms of expectations of v_{t+1}^i rather than $\bar{v}_{t+1}^i$. The IS equation can then alternatively be written in the form

$$Y_t = g_t - \sigma i_t + (1-\beta)b_{t+1} + \int \widehat{E}_t^i v_{t+1}^i di, \tag{18}$$

where one should recall from Equation 5 that b_{t+1} is known at date t (although the debt matures at $t+1$). In this case, because the endogenous public debt matters for output and inflation determination, we must adjoin Equation 5 to the system that describes TE dynamics under a given specification of inflation.

The complete TE dynamics for the endogenous variables $\{\pi_t, Y_t, b_{t+1}\}$ are then given in the non-Ricardian case by the system consisting of Equations 5, 17, and 18, given expectations $\{v_t^i,\ p_t^{*j}\}$ consistent with Equations 6 and 15. One observes that a larger budget deficit (or smaller surplus) should increase output and inflation, to the extent that it does not cause a reduction in the final expectational term in Equation 18, of a magnitude as large as the increase in $(1-\beta)b_{t+1}$.

2.5. Rational Expectations Equilibrium

Above I am completely agnostic about the nature of subjective forecasts. Under the RE hypothesis, all agents' probability beliefs are identical and coincide with the probabilities predicted by one's model, given the choices that people make on the basis of those beliefs. Under the hypothesis that all beliefs are identical, we can replace the operators $\left\{\widehat{E}^i[\cdot]\right\}$ and $\left\{\widehat{E}^j[\cdot]\right\}$ by the single expectation operator $\widehat{E}[\cdot]$. And because in this case $v_t^i = v_t$ for all i and $p_t^{*j} = p_t^*$ for all j, Equations 10 and 15 reduce to

$$\bar{v}_t = -\sigma(i_t - \pi_t) + \widehat{E}_t \bar{v}_{t+1}, \tag{19}$$

$$p_t^* = (1-\alpha)^{-1}[\kappa y_t + u_t] + \beta \widehat{E}_t p_{t+1}^*, \tag{20}$$

respectively. If we assume a monetary policy rule (or central-bank reaction function) of the form[13]

[13]This version of the Taylor rule (Taylor 1993) reflects an implicit inflation target of 0. In particular, if all exogenous disturbances are 0 forever, this policy is consistent with the steady state around which the model equations have been log linearized.

$$i_t = \phi_\pi \pi_t + \phi_y y_t + \epsilon_t^i, \tag{21}$$

where ϵ_t^i is an exogenous disturbance to monetary policy, and use Equations 8, 11, and 21 to substitute for i_t, π_t, and y_t in Equations 19 and 20, we obtain a system that can be written in the form

$$z_t = B\,\widehat{E}_t z_{t+1} + b\xi_t, \tag{22}$$

where z_t is the vector of endogenous variables $(\overline{v}_t,\ p_t^*)$, ξ_t is the vector of (composite) exogenous disturbances $(g_t - Y_t^n,\ u_t,\ \epsilon_t^i)$, and B and b are matrices of coefficients.

If subjective probabilities must coincide with objective probabilities, we can replace Equation 22 with

$$z_t = B\,E_t z_{t+1} + b\xi_t, \tag{23}$$

where $E_t[\cdot]$ denotes an expectation conditional on the state of the world at date t, under the probability distribution over future paths that represents the equilibrium outcome. An REE is then a stochastic process $\{z_t\}$ consistent with Equation 23. (Note that any solution for the process $\{z_t\}$ completely determines the stochastic evolution of the variables $\{\pi_t, Y_t, i_t\}$, using Equations 8, 11, and 21.) We here restrict attention only to the possibility of bounded solutions $\{z_t\}$ on the assumption that the disturbances $\{\xi_t\}$ are bounded, as our log-linearized equations are derived under this assumption.

The RE hypothesis does not necessarily determine a unique set of model-consistent probability beliefs; because it is a consistency criterion, rather than a hypothesis about how beliefs are formed, it essentially defines a fixed-point problem. RE beliefs are a fixed point of the mapping from possible subjective probabilities into the implied objective probabilities (see Evans & Honkapohja 2001 for further discussion of this T mapping). Such a fixed-point problem may or may not have a solution, and the solution may or may not be unique. In the case of a linear system such as Equation 23, Blanchard & Kahn (1980) establish that the existence and uniqueness of bounded solutions depend on the eigenvalues of the matrix B. Because no elements of z_t are predetermined, REE is determinate (a unique bounded RE solution exists) if and only if both eigenvalues of B lie inside the unit circle. Under the assumption that the response coefficients in Equation 21 satisfy $\phi_\pi, \phi_y \geq 0$, one can show (Woodford 2003, chapter 4) that this condition is satisfied if and only if the response coefficients also satisfy

$$\phi_\pi + \frac{1-\beta}{\kappa}\phi_y > 1, \tag{24}$$

sometimes called the Taylor principle.[14] If monetary policy satisfies Equation 24, Equation 23 can be solved forward for z_t as a linear function of current and expected future values of the exogenous disturbances; in the case in which the exogenous disturbances follow linear-Markovian dynamics, so that $E_t\xi_{t+1} = \Lambda\xi_t$ for some stable matrix Λ, this solution is given by

$$z_t = Z\xi_t, \tag{25}$$

where[15]

[14]Readers are referred to Woodford (2003, chapter 4, proposition 4.3). Woodford analyzes a system of the form of Equation 23 but in which the vector z_t has inflation and the output gap as elements. But because this vector is a nonsingular linear transformation of the vector z_t used here (plus an exogenous term, which does not affect the determinacy calculation), the eigenvalues of the matrix B in Woodford (2003) are the same as those of the matrix B here.

[15]Note that this infinite sum must converge because we have assumed that both B and Λ have all eigenvalues inside the unit circle.

$$Z \equiv \sum_{j=0}^{\infty} B^j b \Lambda^j.$$

The implied responses of inflation, output, and interest rates to exogenous disturbances of various kinds are discussed further in Woodford (2003, chapter 4) and Gali (2008, chapter 3).

If instead Equation 24 is not satisfied, there are an infinite number of REE (even restricting our attention to bounded solutions), including solutions in which inflation and output fluctuate in response to sunspots (random events with no consequences for the economic fundamentals ξ_t) or in which the fluctuations in inflation and output are arbitrarily large relative to the magnitude of the exogenous disturbances. In the case of such a policy, the economy may be vulnerable to instability purely because of the volatility of expectations, even under the assumption that the economy must evolve in accordance with an REE. Because instability of this kind is undesirable, it is often argued that a policy commitment should be chosen that ensures the existence of a determinate REE (see, e.g., Woodford 2003, chapters 2 and 4); in the context of NK models of the kind sketched here, this provides an argument for the desirability of an interest-rate rule that conforms to the Taylor principle (see, e.g., Clarida et al. 2000).[16] Apart from the use of the determinacy result as a criterion for the choice of a monetary policy rule, the predicted character of the fluctuations due to self-fulfilling expectations when the Taylor principle is not satisfied has been proposed by some as a positive theory of the aggregate fluctuations observed during periods when monetary policy has arguably been relatively passive (e.g., Clarida et al. 2000, Lubik & Schorfheide 2004). As discussed below, however, relaxation of the RE hypothesis opens up additional possibilities for instability under regimes that fail to pin down expectations sufficiently precisely.

Some argue that avoidance of the indeterminacy of REE need not be a concern when choosing a monetary policy rule, on the grounds that even in the indeterminate case, there is no reason to expect people's expectations to coordinate on a sunspot equilibrium, or even on one with excessive fluctuations in response to fundamental disturbances (e.g., McCallum 1983). Such authors argue that a model's positive prediction should be based on some further refinement of the set of equilibria, such as a restriction to Markovian equilibria, in which endogenous variables depend only on those aspects of the state of the world that affect either the equilibrium relations that determine those variables or the conditional probabilities of states that will be relevant for equilibrium determination in the future.[17]

If $\{\xi_t\}$ is Markovian in the above example, this would mean restricting attention to REE of the form of Equation 25, for some matrix Z. Because Equation 25 represents an REE if and only if the matrix Z satisfies

$$Z = BZ\Lambda + b,$$

and these constitute a system of six linear equations for the six elements of Z, there is a unique solution of this form for generic parameter values, even when the monetary policy rule fails to satisfy Equation 24. [McCallum (1983) calls this the "minimum-state-variable solution."] But

[16]Some object to this argument for the Taylor principle on the grounds that it ensures only a locally unique REE—there is only one equilibrium in which the endogenous variables remain forever near the target steady state—but does not exclude the possibility of other REE, including sunspot equilibria, that do not remain near the steady state (e.g., Benhabib et al. 2001). This issue is beyond the scope of the current review, owing to our reliance on a local log-linear characterization of equilibrium dynamics (but see Woodford 2003, chapter 2, section 4, for further discussion).

[17]This refinement is closely related to the idea of restricting attention to the Markov perfect equilibria of dynamic games (Maskin & Tirole 2001).

the question of whether, or under what circumstances, we should expect people to coordinate on the particular expectations specified by the Markovian solution is one that cannot be answered by the RE hypothesis itself, and the consideration of plausible restrictions on expectations that do not simply assume RE can help justify a particular selection from among the set of REE, as proposed by McCallum.

3. RATIONALIZABLE TEMPORARY EQUILIBRIUM DYNAMICS

As discussed in Section 1, one broad approach to the formulation of a criterion for reasonableness of beliefs—without simply postulating an exact correspondence between people's forecasts and those that are correct (according to one's model of the economy)—is to assume that people should correctly understand the economic model and be able to form correct inferences from it about possible future outcomes. This approach supposes that beliefs should be refined through a process of reflection, independent of experience and not necessarily occurring in real time, that Guesnerie (1992) calls eduction. Although RE beliefs would certainly withstand a process of scrutiny of this kind, such beliefs need not be the only ones that could be rationalized in this way.

3.1. Eductive Stability Analysis

For the sake of concreteness, let us consider the case of Ricardian expectations and monetary policy specified by Equation 21. The assumption that people in the economy understand the model means, in the present context, that people understand that the TE dynamics of inflation, output, and interest rates will be determined by Equations 16, 17, and 21 each period, given the average one-period-ahead forecasts of others. For someone's expectations regarding the paths of the endogenous variables to be consistent with this knowledge, the expected paths must be able to be generated by these equations, under some supposition about the average expectations of others.

Let a possible conjecture about the evolution of average one-period-ahead forecasts be specified by a vector stochastic process $\mathbf{e} \equiv \{e_t\}$, where at any date the two elements of e_t specify $\int \widehat{E}_t^i \overline{v}_{t+1}^i di$ and $\int \widehat{E}_t^j p_{t+1}^{*j} dj$. For any such evolution of average forecasts, the structural equations determine unique TE processes $\{\pi_t, Y_t, i_t\}$. Hence any agent (household or firm) who expects average expectations to evolve in the future in accordance with the process $\mathbf{e}$, and who understands the model, must (to have internally consistent beliefs) forecast precisely this particular evolution of the variables $\{\pi_t, Y_t, i_t\}$. There is then a unique internally consistent anticipated evolution of this agent's own one-period-ahead forecasts $\left\{\widehat{E}_t^i \overline{v}_{t+1}^i\right\}$ (in the case of a household), implied by Equation 10, and similarly a unique internally consistent anticipated evolution $\left\{\widehat{E}_t^j p_{t+1}^{*j}\right\}$ (in the case of a firm), implied by Equation 15. This allows us to determine a new vector stochastic process $\mathbf{e}' \equiv \{e_t'\}$ that describes the one-period-ahead forecasts that must be made by agents who understand the model and believe that the average forecasts of others will evolve according to $\mathbf{e}$.

Let Ψ denote the mapping that determines $\mathbf{e}' = \Psi(\mathbf{e})$ in the way just described. Then individual beliefs $\mathbf{e}^i$ are consistent with knowledge of the model only if there exists some specification of average beliefs $\mathbf{e}$ such that $\mathbf{e}^i = \Psi(\mathbf{e})$. In this case we can say that the beliefs $\mathbf{e}^i$ can be rationalized by the conjecture $\mathbf{e}$ about average beliefs.

Because of the linearity of the mapping Ψ, it is evident that if all agents understand the model, a specification of average beliefs $\mathbf{e}$ is rationalizable if and only if there exists some conjecture about average beliefs $\mathbf{e}^1$ such that $\mathbf{e} = \Psi(\mathbf{e}^1)$. But if additionally all agents understand that all agents understand the model, their conjectures about average beliefs are consistent with this knowledge only if $\mathbf{e}^1$ can itself be similarly rationalized, in other words, if there exists a conjecture $\mathbf{e}^2$ such that $\mathbf{e}^1 = \Psi(\mathbf{e}^2)$. Any number of levels of rationalization might be demanded in a similar way.

Even if we assume that agents' forecasts should be grounded in reasoning of this kind, it may be reasonable only to demand some finite number of levels of rationalization, either because it is assumed only that people understand that others understand . . . that others understand the model, to some finite order of recursion, or because it is not considered practical for agents to check their beliefs for this degree of internal consistency beyond some finite number of levels.[18] In this case, k-th-order beliefs (for some finite k) are allowed to be specified arbitrarily (required to be internally consistent but not necessarily consistent with understanding the model). In this case, obtaining definite conclusions requires a specific theory of k-th-order beliefs (perhaps some fairly simple specification) or at least some bounds on the class of possible specifications of k-th-order beliefs that may be entertained.

Alternatively, as in the literature on rationalizable equilibria in game theory (Bernheim 1984, Pearce 1984), one may require that beliefs be consistent with an infinite hierarchy of beliefs, each level of which is rationalized by the next higher level of beliefs. RE beliefs represent one possible type of rationalizable beliefs in this sense, but not all rationalizable beliefs need be RE beliefs, and even when REE is determinate, there may be a large multiplicity of rationalizable beliefs, even under the requirement that beliefs satisfy some uniform bound at all levels. Guesnerie (1992, 2005) calls an investigation of whether the REE beliefs are the unique rationalizable beliefs "eductive stability analysis." If the REE is eductively stable, he considers the REE outcome to be a reasonable prediction of one's model, but if not, the other rationalizable paths are taken to be equally plausible predictions.

The existence of a large set of possible equilibrium outcomes, including the possibility of fluctuations in response to sunspots or large fluctuations in response to small changes in fundamentals, is regarded as an undesirable form of instability. Hence Guesnerie proposes as a criterion for policy choice the desirability of finding a policy under which the REE is eductively stable.[19] This is in the spirit of the proposal, discussed above, that policy be designed to ensure determinacy of REE, but it is an even stronger requirement as the uniqueness of rationalizable equilibrium necessarily implies the determinacy of REE, while the converse is not true.

3.2. The Taylor Principle and Determinacy Reconsidered

As an illustration, let us consider whether a monetary policy rule of the form of Equation 21 ensures unique (uniformly bounded) rationalizable dynamics. To simplify, as in Guesnerie (2009), let us consider the limiting case of perfectly flexible prices. In this limit, output Y_t is determined by exogenous fundamentals, and Equations 16 and 21 jointly determine i_t and π_t given expectations and exogenous fundamentals. Calculations are also simplified if there are assumed to be no exogenous disturbances (including no variation in Y_t), as the issue of the multiplicity of solutions is unaffected by the amplitude of disturbances. Hence the monetary policy rule given in Equation 21 reduces simply to $i_t = \phi_\pi \pi_t$. REE is determinate if and only if $\phi_\pi > 1$.[20] When this condition is satisfied, the unique bounded REE has $\bar{v}_t = 0$ for all t, implying that $i_t = \pi_t = 0$ for all t.

[18]Readers are referred to, for example, Phelps (1983) and Evans & Ramey (1992) for proposals of this kind. Evans & Ramey propose to endogenize the number of levels of rationalization in terms of "calculation costs" involved in iterating the mapping Ψ another time.

[19]Because the REE is necessarily a rationalizable equilibrium, the uniqueness of rationalizable equilibrium requires that the REE be the only such equilibrium.

[20]This is the implication of Equation 24 in the limit in which $\kappa \to \infty$.

Although $\phi_\pi > 1$ is therefore also a necessary condition for uniqueness of the rationalizable dynamics, it is not sufficient. Guesnerie (2009) shows that if $1/2 < \beta < 1$ and

$$\phi_\pi > (2\beta - 1)^{-1} > 1, \tag{26}$$

then even though REE is determinate, there is a large multiplicity of uniformly bounded rationalizable equilibria. For example, consider the hierarchy of beliefs that may support a rationalizable TE at some date t. The requirement of rationalizability does not establish any necessary linkages between what happens at different dates: only between what happens at date t and what people expect at date t that people will expect at later dates about what people will expect . . . will happen at still later dates. Thus for each date t, we may separately specify what happens at that date, along with the hierarchy of forecasts that rationalize it.

Substituting the policy rule for i_t in Equation 10, and then using Equation 8 to eliminate π_t, we obtain the requirement that

$$\overline{v}_t^i = (1 - \beta\phi_\pi)\overline{v}_t + \beta\widehat{E}_t^i\,\overline{v}_{t+1}^i \tag{27}$$

for all i. In a rationalizable TE, not only must this hold at date t, but everyone must expect anyone else to expect anyone else . . . to expect it to hold at any future date.

One such specification of the hierarchy of beliefs is given by $\overline{v}_t^i = \epsilon,$ and

$$\widehat{E}_t^{i_1}\widehat{E}_{t_1}^{i_2}\cdots\widehat{E}_{t+j_{n-1}}^{i_n}\overline{v}_{t+j_n}^{i_n} = (-\mu)^{1-n}\phi_\pi\epsilon,$$
$$\widehat{E}_t^{i_1}\widehat{E}_{t_1}^{i_2}\cdots\widehat{E}_{t+j_{n-1}}^{i_n}\overline{v}_{t+j_n}^{i_{n+1}} = (-\mu)^{-n}\epsilon,$$

for any sequences of households and dates of the kind assumed above, where ϵ is an arbitrary real number and

$$\mu \equiv \frac{\beta\phi_\pi - 1}{(1-\beta)\phi_\pi}.$$

These beliefs satisfy all the requirements for rationalizability for any real number ϵ, as discussed further in the **Supplemental Appendix** (follow the Supplemental Material link from the Annual Reviews home page at **http://www.annualreviews.org**). If Equation 26 is satisfied, $\mu > 1$, and forecasts of all orders also satisfy a uniform bound. There is thus (at least) a continuum of uniformly bounded rationalizable TE. Moreover, ϵ may represent the realization of a sunspot event unrelated to fundamentals so that there are seen to exist bounded sunspot equilibria, despite the fact that monetary policy satisfies the Taylor principle.

Supplemental Material

If instead

$$1 < \phi_\pi < (2\beta - 1)^{-1}, \tag{28}$$

one can show that the determinate REE is also the only uniformly bounded rationalizable TE. Note that Equation 27 implies that

$$\overline{v}_t = \phi_\pi^{-1}\int \widehat{E}_t^i\overline{v}_{t+1}^i\,di. \tag{29}$$

Hence if it is common knowledge that there exists some finite bound κ such that $|\overline{v}_t^i| \le \kappa$ for all i and t, it follows from Equation 29 that it must also be common knowledge that $|\overline{v}_t| \le \phi_\pi^{-1}\kappa$ for all t. Using this bound, Equation 27 then implies that

$$|\overline{v}_t^i| \leq |1-\beta\phi_\pi|\phi_\pi^{-1}\kappa + \beta\kappa = \lambda\kappa,$$

where

$$0 < \lambda \equiv \max\{\phi_\pi^{-1}, 2\beta - \phi_\pi^{-1}\} < 1.$$

Hence common knowledge that $|\overline{v}_t^i| \leq \kappa$ implies that it must be common knowledge that $|\overline{v}_t^i| \leq \lambda\kappa$. By the same reasoning, it must then be common knowledge that $|\overline{v}_t^i| \leq \lambda^2\kappa$, and so on, until a bound smaller than any positive quantity is established. Thus it must be common knowledge that $v_t^i = 0$ for all i and all t and hence that $i_t = \pi_t = 0$ for all t.

In such a case, Guesnerie states that the REE is "eductively stable" and argues that there is reason (in this case, and this case only) to expect this equilibrium to obtain. Although this is possible, the restrictions on the coefficient ϕ_π are much more stringent than those required for determinacy under the RE hypothesis. If, for example, $\beta = 0.99$ (a common calibration for quarterly NK models), Equation 28 requires that $1 < \phi_\pi < 1.02$. This very tight bound is violated by the rule recommended by Taylor (1993), as well as by most estimated central-bank reaction functions.

If the avoidance of instability due to self-fulfilling expectations of this particular type is a design criterion for monetary policy, it follows that one must be careful about seeking to stabilize inflation in the face of real disturbances simply by using a rule of the form of Equation 21 with a very strong inflation response coefficient. Instead, the simultaneous achievement of eductive stability and stable inflation despite exogenous disturbances is possible only in the case of a policy rule that directly responds to the underlying determinants of inflation, namely, the exogenous disturbances and observed subjective forecasts.

4. LEARNING DYNAMICS

An alternative way of disciplining the specification of expectations does not demand that they be consistent with a correct structural model of the variables that are forecasted but instead requires that the probabilities assigned to possible future outcomes are not too different from the probabilities with which outcomes actually occur. The idea is that it is not reasonable to suppose that people should fail to notice predictable regularities in economic data, regardless of whether they understand why those regularities exist.

But which regularities are the ones that one can reasonably expect people to take into account? A common answer assumes that forecasts should be derived through extrapolation from prior observations. Such approaches, based on explicit models of learning, have the advantage of explaining how the postulated similarity between subjective beliefs and actual patterns in the data comes about. They also reduce the problem of indeterminacy of the model's predictions, pervasive in the case of approaches such as those discussed above, which demand only that beliefs be a fixed point of a certain mapping. Although there may be many possible asymptotic states of belief under an explicit model of learning, with the one that is reached depending on initial conditions and/or random events along the way, a model of learning often makes a unique prediction conditional upon initial conditions and the subsequent history of shocks.

The most common approach of this general type assumes that agents' forecasts at any time t are derived from an econometric model, estimated using the data observed up until that date.[21] Let the

[21]Readers are referred to Evans & Honkapohja (2001, 2009) and Sargent (2008) for reviews of work of this kind.

model be specified by a vector θ of parameters and suppose that any model θ implies that forecasts e_t should be some function of the current state ζ_t. Then, in any period t, new estimates $\widehat{\theta}_t$ and $\widehat{\zeta}_t$ are formed of the parameters and of the state, based on the data available at that point. Under a recursive estimation scheme, the new estimates are functions of the prior estimates and of the new data observed since the formation of the prior estimates,

$$\widehat{\theta}_t = \Theta_t\left(\widehat{\theta}_{t-1}, \zeta_{t-1}, x_t\right), \tag{30}$$

$$\widehat{\zeta}_t = Z_t\left(\widehat{\theta}_{t-1}, \widehat{\zeta}_{t-1}, x_t\right), \tag{31}$$

where x_t is some vector of new data.[22] Given these new estimates, period t forecasts are given by a function[23]

$$e_t = \Psi\left(\widehat{\zeta}_t; \widehat{\theta}_t\right). \tag{32}$$

Given these beliefs, the TE values of the variables z_t are determined by Equation 22, given the exogenous disturbances ξ_t. This system, possibly along with additional structural equations, determines the new data x_t.[24] Thus the system of Equations 22 and 30–32 jointly determines $\widehat{\theta}_t, \widehat{\zeta}_t, e_t$, and x_t, given the lagged estimates and the disturbances ξ_t. Solution of these equations in each of a succession of periods yields the predicted dynamics of both beliefs and endogenous variables, as a function of the history of exogenous disturbances.

4.1. Restricted Perceptions Equilibrium

A focus of much of the literature on TE dynamics with learning has been to ask whether learning dynamics should converge asymptotically to REE; indeed, much of the early literature (beginning with Bray 1982) was concerned more with the foundations of the REE concept—seeking to provide a causal explanation for how the postulated coincidence between subjective and objective probabilities could come about—than with the provision of an alternative model of economic dynamics. Obviously this is only possible if the class of forecasting models that are contemplated includes a model θ^{RE} that produces the forecasts associated with the REE. If one does not assume that economic agents are endowed with knowledge of the structural model, and hence with the information required to compute the REE, it is not obvious that their forecasting approach should

[22]The time subscripts on the functions $\Theta_t(\cdot)$ and $Z_t(\cdot)$ allow for the possibility that the updating rules may depend on the size of the existing data set. Equation 44 for the evolution of the mean estimates provides a simple example.

[23]Here, for simplicity, I assume that every household i and every firm j forecasts in the same way, using the same observed data, so that average forecasts are simply the common forecasts, given by a function of the common estimates. One may, however, allow each household to have its own estimated model $\widehat{\theta}^i_t$, evolving according to a separate equation of the form of Equation 30, and then define $\int \widehat{E}^i_t \overline{v}^i_{t+1} di$ as a function of the entire probability distribution of estimates $\left\{\widehat{\theta}^i_t\right\}$, rather than simply as a function of a single estimate $\widehat{\theta}_t$, and similarly with the firms.

[24]It is common in the literature on learning dynamics to specify a recursive causal structure by assuming that the data x_t are actually determined in period $t-1$ (i.e., they include π_{t-1} rather than π_t). In this case, all the arguments of the functions in Equations 30 and 31 are predetermined so that these equations determine the new estimates $\left(\widehat{\theta}_t, \widehat{\zeta}_t\right)$ independently of the period t shocks; Equation 32 then determines the forecasts e_t, and finally Equation 22 determines the endogenous variables z_t given the shocks. But it is not obvious why, in the logic of the NK model presented here, one should suppose that period t forecasts must be made prior to the observation of period t endogenous variables.

even entertain as a possibility the precise forecasting rule implied by the REE, but if no value of θ results in forecasts of this kind, convergence to REE beliefs (and hence to the REE dynamics) is obviously impossible.

There might, however, still be convergence of beliefs to some fixed point $\bar{\theta}$ with the property that under the TE dynamics generated by the beliefs $\bar{\theta}, \bar{\theta}$ is the model (among the class of models considered) that yields the best forecasts (under some criterion that is used for the estimation). For example, suppose that the class of models considered consists of those in which both $\bar{v}_{t+1}$ and p^*_{t+1} are linear functions of ζ_t and unforecastable disturbances at $t+1$, that the elements of ζ_t are all part of the history of the observables $\{x_t, x_{t-1}, \ldots\}$ so that ζ_t is observable, and that the coefficients of the linear model are estimated so as to minimize the mean squared error of the forecasts of $\bar{v}_{t+1}$ and p^*_{t+1}. Then forecasts will be of the form $e_t = \widehat{\theta}'\zeta_t$, where the vector ζ_t is assumed to include an element equal to 1 each period, and with an infinite sequence of data generated by the TE dynamics under beliefs $\bar{\theta}$, the estimated coefficients will satisfy

$$\widehat{\theta}' = \mathrm{E}\left[z_{t+1}\zeta_t'\right]\mathrm{E}\left[\zeta_t\zeta_t'\right]^{-1}, \tag{33}$$

where E[·] indicates an unconditional expectation under the ergodic TE dynamics resulting from some beliefs θ. Alternatively, we can write $e_t = P_t[z_{t+1}]$, where $P_t[\cdot]$ denotes the linear projection of the random variable inside the brackets on the space spanned by the elements of ζ_t.

The beliefs $\bar{\theta}$ constitute a restricted perceptions equilibrium (RPE) if the optimal estimate $\widehat{\theta}$ given by Equation 33 is equal to $\bar{\theta}$ when the unconditional expectations are the ones implied by the TE dynamics generated by beliefs $\theta = \bar{\theta}$ (Evans & Honkapohja 2001, chapter 13; Branch 2004). This is a weaker requirement than that of an REE, as forecasts are assumed to be optimal only within a particular class of linear models, rather than within the class of all forecasts that might be made on the basis of information available in period t. Note that in the special case in which the optimal forecast of z_{t+1} is indeed a linear function of ζ_t so that

$$E_t[z_{t+1}] = P_t[z_{t+1}] = T(\theta)'\zeta_t \tag{34}$$

when the TE dynamics are generated by beliefs θ, then Equation 33 implies that $\widehat{\theta} = T(\theta)$. Hence in this case, $\bar{\theta}$ describes RPE beliefs if and only if $T(\bar{\theta}) = \bar{\theta}$, which is also the condition for REE beliefs. More generally, however, when the forecasting variables ζ_t do not span a large-enough space, or at any rate not the correct one, RPE beliefs will differ from REE beliefs.

Conditions can be established under which the learning dynamics resulting from repeated re-estimation of an ordinary least-squares forecasting equation (least-squares learning dynamics) converge with probability 1 to an RPE as the length of the observed data set grows large enough. But even in such a case, the dynamics need not coincide, even asymptotically, with the model's REE dynamics. Fuster et al. (2010, 2011, 2012) provide examples in which more complex dynamics of asset prices, consumption, and investment are implied by RPE dynamics than would be associated with REE dynamics of the same models; here the suboptimality of forecasts results from estimation of lower-order autoregressive models of the data than the correct model.

4.1.1. Application to the new Keynesian model. Suppose that we equate the subjective expectations in Equation 22 with linear projections to obtain

$$z_t = BP_t z_{t+1} + b\xi_t. \tag{35}$$

This is the set of conditions that must be satisfied for the evolution of the expectational variables $\{z_t\}$ to represent an RPE, under either of two possible interpretations of how forecasts for horizons more than one period in the future are formed.

On the one hand, we might assume, as Preston (2005) does, that forecasts for arbitrary future horizons are formed by estimating a vector-autoregressive system

$$P_t x_{t+1} = \Lambda_x \zeta_t, P_t \zeta_{t+1} = \Lambda_\zeta \zeta_t,$$

where x_t is the vector of variables that must be forecasted, other than the future values of the forecasting variables ζ_t themselves. Forecasts for arbitrary future horizons can then be computed as

$$\widehat{E}_t^i x_{t+j} = \widehat{E}_t^i \widehat{E}_{t+1}^i \cdots \widehat{E}_{t+j-1}^i x_{t+j} = \Lambda_x \Lambda_\zeta^{j-1} \zeta_t$$

for any $j \geq 1$. That is, forecasts for horizons more than one period in the future are formed by forecasting one's own future one-period-ahead forecasts, whereas one-period-ahead forecasts are given by linear projections on the forecasting variables ζ_t.[25] Given forecasts of this kind, the definitions in Equations 9 and 12 of the expectational variables then imply that their evolution must satisfy Equation 35.

On the other hand, we might assume, as Evans & McGough (2009) propose, that people estimate the values of the expectational variables z_t not using the definitions in Equations 9 and 12 of these variables in terms of long-horizon forecasts of variables with objective definitions, but instead using the recursive relations given in Equations 10 and 15 to estimate values on the basis of one's current forecasts of one's own estimates in the next period. These forecasts of one's own future estimates can be obtained by collecting data on what one's estimates have been and regressing them on the previous period's values of the forecasting variables ζ_t. Also under this assumption, if there is convergence to an RPE, the expectational variables will necessarily satisfy Equation 35. Note that these two approaches to least-squares learning are not mathematically equivalent, but if in each case there is convergence to an RPE, then the RPE is the same in both cases.[26]

Given a solution for the dynamics of the expectational variables $\{z_t\}$, the dynamics of inflation, output, and interest rates are then determined by Equations 8, 11, and 21, as under other specifications of expectations. The difference between REE and RPE predictions then stems entirely from the difference between Equations 23 and 35. When Equation 34 holds, the predictions are necessarily the same.

4.1.2. Failure of Ricardian equivalence. As an illustration of how macroeconomic dynamics in an RPE may differ from the REE dynamics predicted in the case of a given policy rule, let us reconsider the argument for Ricardian equivalence. Suppose that expectations are not Ricardian (i.e., that people do not assume that the future path of primary surpluses must satisfy Equation 7) and instead estimate an econometric model to forecast future surpluses that does not impose this condition as an a priori restriction.[27] The TE dynamics are then determined by the system consisting of Equations 5, 17, and 18, given expectations of the future evolution of the variables $\{p_t^{*j}, v_t^i\}$ defined by Equations 6 and 15, and specified paths for the policy variables $\{i_t, s_t\}$.

[25] An advantage of this method is that forecasts $\widehat{E}_t^i x_{t+j}$ can be formed for arbitrarily large j, using coefficients that can be estimated using finite data sets, as there is no need to actually regress observed values of x_{t+j} on the prior forecasting variables ζ_t.

[26] Because the learning dynamics outside the RPE are in general not identical, the conditions for convergence to an RPE are not always the same in the two cases.

[27] Note that each household needs only to forecast its own tax obligations in excess of the value of government purchases; its use of a forecasting model that violates Equation 7 does not necessarily imply that it believes that aggregate tax collections do not satisfy the present-value relation. Although I assume that the tax obligations of all households are the same, this is not necessarily known to the households.

Suppose further that monetary policy is described by a Taylor rule of the form of Equation 21, the coefficients of which satisfy the Taylor principle (Equation 24), while fiscal policy is described by a feedback rule of the form

$$s_t = \phi_b b_t + \epsilon_t^s, \tag{36}$$

where

$$1 - \beta < \phi_b < 1, \tag{37}$$

and ϵ_t^s is an exogenous disturbance. This specification, together with Equation 5, implies debt dynamics that remain bounded in the case of any bounded processes for inflation and the nominal interest rate, and hence that model-consistent expectations will be Ricardian, in the case of any REE involving bounded fluctuations. Such a specification of monetary and fiscal policy implies the existence of a determinate REE in the case of any bounded disturbance processes, and in this REE, the fiscal shocks $\{\epsilon_t^s\}$have no effects on the evolution of output, inflation, or interest rates.[28] Hence in such a model, RE imply Ricardian equivalence.

Let us consider instead the possible character of RPE dynamics. Suppose, for example, that the vector of forecasting variables ζ_t consists only of the single state variable s_t.[29] Then RPE forecasts are of the form $e_t = \psi s_t$,[30] for some vector of coefficients ψ. Substituting these forecasts into Equations 17 and 18, one can solve for the TE values of π_t, y_t, i_t, and b_{t+1} as linear functions of b_t, s_t, and ξ_t.

The calculations are especially simple if we consider a case in which $\phi_y = s_b = \kappa = 0$ and assume that there are no exogenous disturbances other than the fiscal shock $\{\epsilon_t^s\}$, which is assumed to be unforecastable (white noise). In this limiting case, there are no equilibrium fluctuations in π_t or i_t, and the solutions for y_t and b_{t+1} are given by

$$y_t = \left(\beta^{-1} - 1\right)(b_t - s_t) + \psi_v s_t, \tag{38}$$

$$b_{t+1} = \beta^{-1}(b_t - s_t), \tag{39}$$

where ψ_v (to be determined) is the first element of ψ. It then follows from Equation 6 that

$$v_t^i = y_t - (1 - \beta) b_t \tag{40}$$

for all i.

From this one can show that

$$\begin{aligned} e_{1t} &= P_t\left[y_{t+1} - (1-\beta) b_{t+1}\right] \\ &= \left(\beta^{-1} - 1\right)(1-\beta) P_t[b_{t+1}] + \left(\psi_v + 1 - \beta^{-1}\right) P_t[s_{t+1}] \\ &= \left[(1 - \beta - \phi_b)\left(\beta^{-1} - 1\right) + \phi_b \psi_v\right] P_t[b_{t+1}], \end{aligned}$$

[28]Readers are referred to Woodford (2001) for further discussion of the consequences of a locally Ricardian fiscal policy of this kind, under an REE analysis.

[29]Note that under the REE dynamics, if the disturbances are all unforecastable (white noise), none of the state variables that must be forecasted by households or firms is forecastable, except the primary surplus (which must be forecasted to estimate v_t^i). It is perhaps not implausible to suppose that households forecast future primary surpluses using only the current level of the surplus. This would not, however, constitute a model-consistent forecast, as Equations 5 and 36 imply that an optimal forecast of future primary surpluses depends on the value of b_{t+1} or alternatively upon both s_t and b_t.

[30]In this non-Ricardian case, the first element of e_t is assumed to be $\int \hat{E}_t^i v_{t+1}^i di$ rather than $\int \hat{E}_t^i \bar{v}_{t+1}^i di$.

where for any variable x_{t+1} known at date t, $P_t[x_{t+1}]$ denotes the linear projection of x_{t+1} on s_t. (Here the first line uses the fact that Equation 40 must also hold at date $t+1$; the second line uses the fact that Equation 38 must also hold at $t+1$; and the third line uses the fact that s_{t+1} is determined by Equation 36.) Writing $P_t[b_{t+1}] = \Lambda_b s_t$, where the coefficient Λ_b depends only on β and ϕ_b (not the assumed value of ψ_v), one then observes that ψ_v must satisfy the consistency condition

$$\psi_v = \left[(1-\beta-\phi_b)\left(\beta^{-1}-1\right)+\phi_b\psi_v\right]\Lambda_b. \tag{41}$$

Under the assumption given in Equation 37, this equation has a unique solution for ψ_v and implies that[31]

$$\psi_v < \beta^{-1} - 1. \tag{42}$$

Equation 38 then implies that in the unique RPE, an exogenous positive innovation in the size of the primary surplus s_t lowers current output y_t. It will also reduce the debt b_{t+1} carried into the next period, with persistent effects on economic activity in later periods as well. Hence Ricardian equivalence does not hold in the RPE, even though the specification of fiscal policy implies that in the model's unique bounded REE, fiscal shocks have no effect on output, either immediately or subsequently. Although Equation 39 implies that even under the RPE dynamics, a correct forecast would satisfy Equation 7 at all times, households' forecasts do not satisfy this condition, as a result of forecasting future surpluses purely on the basis of the current primary surplus, and because of this systematic forecasting error, Ricardian equivalence fails.

4.2. Learnability of Rational Expectations Equilibrium

Even when the class of contemplated forecasting models does include the REE forecasts, and even when the estimator used to determine $\widehat{\theta}$ is one that should be asymptotically consistent, in the case of a sufficiently long series of data generated by the REE, it need not follow that $\widehat{\theta}$ must converge asymptotically to θ^{RE} under the TE dynamics with learning. The reason is that, at each point in time, the observed data will actually be generated by the behavior that results from current beliefs $\widehat{\theta}_t$, and not by REE behavior. If a departure of people's estimates from θ^{RE} gives rise to patterns in the data that justify estimates even further from θ^{RE}, the learning dynamics may diverge from RE beliefs almost surely, even if people start out with beliefs quite near to RE beliefs. The question of whether the REE can in fact be reached as the asymptotic outcome of a learning process of the kind described above is therefore a nontrivial one. Authors such as Bullard & Mitra (2002) and Evans & Honkapohja (2003, 2006) propose as a design criterion for a monetary policy rule not only that the rule should be consistent with a desirable REE, but that the rule should imply that learning dynamics should converge to that REE so that the desirable equilibrium is learnable.

4.2.1. Adaptive estimation of means. As a simple example, suppose that the disturbances ξ_t are all independently and identically distributed (i.i.d.) random variables, with mean 0. In this case, there is a unique Markovian REE, in which $z_t = b\xi_t$ each period, and the RE forecasts satisfy $E_t z_{t+1} = 0$ at all times. In this equilibrium, π_t, y_t, and i_t will also each be a linear function of ξ_t, and the RE forecast of each variable will be 0 (i.e., the constant steady-state value) at all times. Suppose, furthermore, that the class of forecasting models considered by decision makers consists of all models under which the forecast of each variable is a constant (that is, people believe that each of these is an i.i.d. random variable and seek only to estimate its mean). This simple class

[31]Readers are referred to the **Supplemental Appendix** for details.

of models includes the forecasting rule used in the Markovian REE, so the assumption of such a restricted class does not in itself rule out the possibility of convergence to RE beliefs.

Finally, suppose that people estimate the means of each of the stationary variables using the sample mean of the values observed to date so that

$$\widehat{x}_t = t^{-1} \sum_{s=1}^{t} x_s, \tag{43}$$

where x_t refers to any of the variables π_t, y_t, and i_t or to any of the elements of ξ_t; $\widehat{x}_t$ is the estimate of the variable's mean at date t (common to all agents); and 1 is the date at which the available data series begins.[32] This can be written recursively in the form

$$\widehat{x}_t = \widehat{x}_{t-1} + \gamma_t \left(x_t - \widehat{x}_{t-1} \right), \tag{44}$$

where the gain $\gamma_t = 1/t$ indicates the degree to which estimates are adjusted in response to an observation that differs from what has been forecasted. Note that in the case in which the economy were to reach the REE, each of the variables x_t would indeed be i.i.d., and the estimators in Equation 43 would almost surely converge asymptotically to the true means (and hence to the REE beliefs) by the law of large numbers. Hence the estimation strategy is not inherently incompatible with learning the REE beliefs.

Any set of estimates of the means implies forecasts given by[33]

$$\widehat{E}_t^i \overline{v}_{t+1}^i = \widehat{y}_t - \widehat{g}_t - \frac{\sigma}{1-\beta} \left(\beta \widehat{i}_t - \widehat{\pi}_t \right), \tag{45}$$

$$\widehat{E}_t^j p_{t+1}^{*j} = \frac{1}{1-\alpha\beta} \widehat{\pi}_t + \xi \widehat{y}_t + \widehat{\mu}_t, \tag{46}$$

for each household and each firm. Hence in vector form we can write

$$e_t = C \widehat{\mathbf{x}}_t, \tag{47}$$

for a certain matrix of coefficients C, where $\widehat{\mathbf{x}}_t$ is the vector of estimates in Equation 43. The system consisting of Equations 16, 17, and 21 allows us to solve for TE values

$$\mathbf{x}_t = D e_t + d \xi_t, \tag{48}$$

for certain matrices D and d, where $\mathbf{x}_t$ is the vector of actual values of the variables x_t. Equations 44, 47, and 48 then completely describe the TE dynamics of actual values and forecasts with adaptive learning, given the exogenous disturbances $\{\xi_t\}$ and initial prior estimates $\widehat{x}_{t-1}$.

Combining these equations, we obtain a law of motion

$$[I - \gamma_t DC] \widehat{\mathbf{x}}_t = (1-\gamma_t) \widehat{\mathbf{x}}_{t-1} + \gamma_t d \xi_t \tag{49}$$

for the estimates; thus the TE dynamics are uniquely defined as long as the matrix in square brackets is nonsingular, as I assume.[34] In fact, all that matters about these estimates is the implied

[32]This is an example of least-squares learning, in which the vector s_t has a single element, 1, each period.

[33]Here I use the notation $\widehat{g}_t$ for the current estimate of the mean of the composite disturbance $g_\tau - Y_\tau^n$ for simplicity.

[34]For any matrices C and D, this will be true for all small-enough values of the gain γ_t. We are here concerned only with TE dynamics in the low-gain case.

forecasts e_t, so we can reduce the dimension of the system in Equation 49 by premultiplying by the matrix C, yielding

$$[I - \gamma_t A]\Delta e_t = -\gamma_t [I - A] e_{t-1} + \gamma_t a \xi_t, \tag{50}$$

where $A \equiv CD$, $a \equiv Cd$. This equation determines the dynamics of the forecasts $\{e_t\}$ given initial forecasts and the evolution of the exogenous disturbances; the paths of the other relevant variables are then given by Equation 48. The TE dynamics converge asymptotically to the REE dynamics (and subjective expectations coincide asymptotically with the REE forecasts) if and only if $e_t \to 0$ for large t.

Using the stochastic approximation methods introduced by Marcet & Sargent (1989) and expounded in detail by Evans & Honkapohja (2001), one can show that in the case of a decreasing gain sequence $\{\gamma_t\}$ similar to the one implied by Equation 43, the path for $\{e_t\}$ implied by the stochastic law of motion given in Equation 50 eventually converges to one of the trajectories of the ordinary differential equation (ODE) system

$$\dot{e} = -[I - A] e(\tau), \tag{51}$$

where τ is a rescaled time variable defined by $\tau_t \equiv \sum_{s=1}^{t} \gamma_s$, and the dot indicates a derivative with respect to τ.

The ODE system given in Equation 51 has a unique rest point $\bar{e} = 0$ (corresponding to REE forecasts) if and only if the 2×2 matrix A has no eigenvalue exactly equal to 1 (so that $I - A$ is nonsingular), and the trajectories of Equation 51 converge asymptotically to this rest point if and only if both eigenvalues of A have real parts less than 1 (so that both eigenvalues of $I - A$ have positive real parts). If the latter condition holds, Equation 50 implies that $e_t \to 0$ with probability 1 as t grows large; beliefs asymptotically approach the REE forecasts, and one may say that the REE is learnable following the procedure postulated above. If, instead, A has an eigenvalue with real parts greater than 1, the trajectories of Equation 51 diverge from the rest point for almost all initial conditions, and correspondingly, one can show that there is zero probability of the beliefs implied by Equation 50 remaining forever within a neighborhood of the REE beliefs, even if people begin with initial beliefs near (or exactly equal to) the REE beliefs.

Hence the learnability of the REE depends on the eigenvalues of A. In the case of a policy (Equation 21) with $\phi_\pi, \phi_y \geq 0$, one can show that both eigenvalues of A have real parts less than 1 (implying learnability) if and only if the response coefficients satisfy Equation 24—that is, policy conforms to the Taylor principle.[35] This is identical to the condition for determinacy of the REE dynamics, and the connection between the two results is not accidental. The matrix A has an eigenvalue equal to 1 if and only if the model's steady-state inflation rate and output gap are indeterminate: The associated right eigenvector $\bar{e}$ indicates the direction in which a constant forecast ($e_t = \bar{e}$ for all t) may differ from 0 and still constitute a perfect-foresight equilibrium if the disturbances equal 0.[36] (Any multiple of $\bar{e}$ is also a possible perfect-foresight steady state in such a case.) But there exists a continuum of steady states if and only if the matrix B in Equation 23 has an eigenvalue equal to 1, and $\bar{e}$ must also be the associated right eigenvector of B. We have seen above that B has such an eigenvalue if and only if Equation 24 holds with equality.

[35]In fact, when Equation 24 is satisfied, both eigenvalues of A are inside the unit circle. When it fails, there is a real eigenvalue greater than 1.

[36]If $A\bar{e} = \bar{e}$, then forecasts $e_t = \bar{e}$ each period lead to outcomes $\mathbf{x}_t = D\bar{e}$ each period. Then $\hat{\mathbf{x}}_t = D\bar{e}$ will be a perfect-foresight estimate each period, implying that the correct forecasts each period will be $CD\bar{e} = \bar{e}$.

Intuitively, the Taylor principle guarantees determinacy of the REE dynamics because perturbations of the expected future values of the elements of z result in a current TE value of z_t that is closer to 0 than whatever is expected for the next period; hence REE forecasts $E_t z_{t+j}$ cannot be bounded for all j and also consistent with this contraction requirement, unless they are exactly 0 for all j. But the fact that forecasts e_t different from 0 give rise to TE values z_t that are closer to 0 (on average) also implies that adaptive learning will move the forecasts closer to 0 (on average) so that the learning dynamics eventually converge to the REE forecasts (see Woodford 2003, chapter 4, section 2.3, for further discussion).

The analysis above assumes a very simple kind of least-squares learning, in which the only contemplated forecasting rules are ones in which the forecasts are constants (estimates of the means of the various variables) and the same for all horizons. But Preston (2005) establishes the same conditions for learnability of the REE when people use forecasting rules of the form

$$\widehat{E}_t x_{t+1} = \widehat{a} + \widehat{b}' \xi_t$$

for each variable x and estimate the coefficients $\widehat{a}$, $\widehat{b}'$ by regressing observations of x_{t-j} on ξ_{t-j-1} (for $0 \leq j \leq t-1$). Longer-horizon forecasts are then formed using

$$\widehat{E}_t x_{t+k} = \widehat{a} + \widehat{b}' \widehat{E}_t \xi_{t+k-1},$$

where forecasts of the future disturbances are based on estimated autoregressive models of each disturbance (see the discussion in Section 4.1.1 for further description of Preston's method of vector-autoregressive-based forecasts).

A policy inconsistent with the Taylor principle still leads to instability because perturbations of the estimated constant terms $\widehat{a}$ result in average values of the variables that differ from 0 by an even greater amount, leading to explosive dynamics for the estimates as above. Conversely, conformity to the Taylor principle remains sufficient for stability of the learning dynamics because the conditions under which estimates of the response coefficients $\widehat{b}'$ diverge are even more restrictive than those required for divergence of the estimates of the constant terms. Preston (2005) also shows that the Taylor principle is necessary and sufficient for learnability of the minimum-state-variable REE (Equation 25) using this approach, in the case in which the disturbances are AR(1) processes (and hence Markovian). This result again follows because the key to the convergence to the REE forecasting rule is the convergence of the estimates of the constant terms $\widehat{a}$, and the mean dynamics of these estimates are unaffected by the stationary fluctuations in the disturbances.[37]

4.2.2. The possibility of a deflation trap. Thus under the least-squares learning approach we again conclude that a rule (Equation 21) that fails to conform to the Taylor principle (Equation 24) makes the economy vulnerable to instability due to self-fulfilling fluctuations, although through a different mechanism than in Section 2.5 above. The explosive dynamics of forecasts in the case of insufficient feedback from aggregate outcomes (especially for inflation) to the interest-rate target generalize the informal argument of Friedman (1968) about the instability resulting from an interest-rate peg.[38]

[37]Bullard & Mitra (2002) reach a similar conclusion, although on the basis of assumed TE dynamics derived by substituting subjective expectations for objective expectations in certain Euler equations of the REE model, rather than deriving the TE dynamics from infinite-horizon optimization under subjective expectations, as above. For comparison of this Euler-equation approach to modeling learning dynamics with the one used here, readers are referred to Preston (2005) and Evans & McGough (2009).

[38]Howitt (1992) presents the first attempt to formalize Friedman's argument through an analysis of the convergence of learning dynamics to the REE.

The problem is not necessarily avoided, however, by commitment to a rule that satisfies the Taylor principle. The reason is that the linear TE dynamics analyzed above cannot hold globally; in particular, policy cannot be described by Equation 21 for all possible inflation rates and output gaps because of the zero lower bound (ZLB) on the nominal interest rate.[39] Even if the central bank follows Equation 21 with coefficients satisfying Equation 24 until the lower bound becomes a binding constraint, the altered response at low levels of inflation and output implies the existence of a second (deflationary) perfect-foresight steady state consistent with the policy rule, as discussed by Benhabib et al. (2001), and the insensitivity of the interest rate to variations in inflation and output once the lower bound binds implies that the learning dynamics will be unstable near the forecasting rule associated with the deflationary REE.[40]

This is illustrated in **Figure 1**, in which the global behavior of the ODE system corresponding to Equation 51 is plotted now taking into account the ZLB. Note that the third row of vector Equation 48 can be written as $i_t = D_i' e_t$, if we average out the values of ξ_t (in order to describe the mean dynamics, which approximately characterize the asymptotic dynamics of our system). The asymptotic dynamics described by Equation 51 are therefore consistent with the ZLB as long as $e(\tau)$ remains in the region satisfying the inequality[41]

$$\bar{r} + D_i' e \geq 0, \tag{52}$$

where $\bar{r} > 0$ is the steady-state real rate of interest.[42] Because both elements of D_i are positive under the sign assumptions stated above, the region in **Figure 1** where the dynamics in Equation 51 apply is the region above and to the right of the line labeled ZLB, along which Equation 52 holds with equality.

When Equation 52 is violated, Equation 21 must instead be replaced by

$$i_t = -\bar{r} < 0. \tag{53}$$

Solving the system consisting of Equations 16, 17, and 53, one obtains a linear solution of the form

$$x_t = \underline{x} + \underline{D}\, e_t + \underline{d}\, \xi_t \tag{54}$$

instead of Equation 48. The complete TE solution for x_t is then given by Equation 48 when Equation 52 is satisfied and Equation 54 otherwise. (Note that this is a continuous, piecewise-linear solution.) Repeating the derivation of Equation 51, one finds that $\dot{e}$ is given by Equation 51 when Equation 52 is satisfied (a region that includes a neighborhood of the origin) and is given instead by

$$\dot{e} = -\left(I - \underline{A}\right)e + C\underline{x} \tag{55}$$

in the region where the inequality is reversed, where $\underline{A} \equiv C\underline{D}$. (Note that this makes $\dot{e}$ a continuous, piecewise-linear function of e.) The trajectories of this system are plotted in **Figure 1**. (Analytical derivations of qualitative properties of this figure are given in the **Supplemental Appendix**.)

Supplemental Material

[39]The other structural relations assumed above are merely local approximations to relations that should actually be nonlinear, but even if they are assumed to hold globally, the zero lower bound (ZLB) prevents Equation 50 from holding globally.

[40]Evans & Honkapohja (2010) and Benhabib et al. (2012) show this in the context of NK models with adaptive learning closely related to the one presented here.

[41]Recall that in the notation used here, i_t is the amount by which the nominal interest rate exceeds its steady-state value so that the requirement for the nominal interest rate to be nonnegative is $\bar{r} + i_t \geq 0$.

[42]Because we log-linearize our equations around a stationary equilibrium with zero inflation, $\bar{r}$ is also the steady-state nominal interest rate.

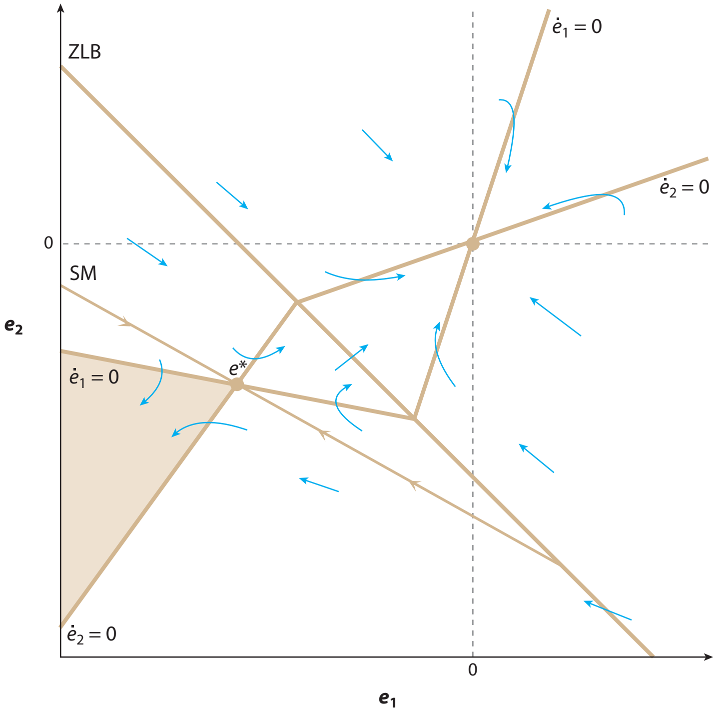

Figure 1

Ordinary differential equation (ODE) trajectories that approximate asymptotic learning dynamics, when interest-rate policy is constrained by the zero lower bound (ZLB). The points at which the constraint binds are those below and to the left of the ZLB line. The steady state in which the inflation target is achieved (corresponding to the origin) is a locally stable rest point of the ODE dynamics, but there is also a second steady state (point e^*) at which the ZLB constraint binds, with stable manifold (SM).

One observes that the origin $e = 0$ (corresponding to the zero-inflation steady state) is a rest point of these dynamics and is locally stable under the ODE dynamics as discussed above. If the dynamics in Equation 51 applied globally (i.e., if the ZLB constraint were not an issue), this steady state would also be globally stable: The learning dynamics would converge to it asymptotically from all possible initial states of belief. But the dynamics when the ZLB constraint binds are different; as a consequence, there is a second steady state in the region below the ZLB line, at

$$e = e^* \equiv \left(I - \underline{A}\right)^{-1} C \underline{x},$$

corresponding to steady-state values

$$\pi^* = i^* = -\bar{r} < 0, \quad y^* = -(1-\beta)\bar{r}/\kappa < 0.$$

Because $I - \underline{A}$ has two real eigenvalues, one positive and one negative, trajectories of Equation 55 converge to e^* only from initial conditions along the SM line in **Figure 1**, the one-dimensional stable manifold. Trajectories above and to the right of this line eventually converge to the zero-inflation steady state, whereas those below and to the left of it diverge from e^* in the opposite direction, eventually being drawn into (and remaining forever in) the shaded region. Because the actual dynamics of inflation are stochastic (even for arbitrarily large t) rather than precisely equal to the approximating ODE dynamics, there is actually zero probability of convergence of the learning dynamics to the REE represented by e^*, even from initial conditions on the SM line; the learning dynamics must diverge from e^* in one direction or the other.[43]

One might think that the nonlearnability of the deflationary REE (while the learning dynamics are instead locally convergent near the REE consistent with the central bank's inflation target) implies that one need not be concerned about the possibility of falling into a self-fulfilling deflation trap of the kind stressed by Benhabib et al. (2001) on the basis of the REE analysis. But the divergent dynamics near the deflationary REE include the existence of trajectories that diverge in the direction of ever-lower levels of inflation and output (those in the shaded region of **Figure 1**),[44] as a result of which the learning dynamics do imply the possibility of a deflation trap, albeit not one that involves convergence to the deflationary REE emphasized by Benhabib et al. As Evans & Honkapohja (2010) and Benhabib et al. (2012) discuss, to the extent that expectations are necessarily formed in this backward-looking way, the only way out of such a trap is to use other tools of policy (such as fiscal stimulus[45]) to raise inflation and/or output long enough for inflation and output expectations to return to the region in which the learning dynamics can be expected to converge toward the target REE without further artificial support. In this view, other tools of policy have an important stabilization role to play in deep crises, even if monetary policy alone suffices as a stabilization tool except when unusual shocks drive expectations far enough away from the target REE forecasting rule.

4.3. Learning Dynamics as a Source of Persistence

Much of the early literature on TE dynamics with learning was concerned with the question of asymptotic convergence to an REE; the positive prediction of interest was whether an REE (or which REE) should be reached, and hence observed in practice. But the learning dynamics themselves might also be regarded as a source of positive predictions. One such positive prediction of particular interest is the existence of persistent fluctuations resulting from the dynamics induced by evolving estimates of the coefficients of people's forecasting rules.

4.3.1. Constant-gain learning. To study the macroeconomic dynamics that result from learning, it is convenient to assume that the gain γ_t in Equation 44 takes some constant value $0 < \gamma < 1$ for all t so that convergence to the REE never occurs, even asymptotically. A constant-gain learning algorithm of this kind may be justified as making sense if people believe that the coefficients of the

[43]Evans & Honkapohja's (2010) figure 2 is qualitatively similar, although plotted in the plane of inflation and output expectations $(\widehat{\pi}_t, \widehat{y}_t)$. These authors obtain an autonomous differential equation system in the $\widehat{\pi}$-$\widehat{y}$ plane only by assuming that interest-rate forecasts are obtained from inflation and output forecasts using people's knowledge of the policy rule. This is not consistent with the assumption made here that interest rates (as with all other variables) are forecasted on the basis of past observations of that variable. In particular, when trajectories in **Figure 1** cross the ZLB line, the nominal interest rate becomes positive, and under the learning rule assumed here, a positive nominal interest rate must be expected in the future as well. Under the Evans-Honkapohja forecasting assumption, instead people would continue for some time to forecast a zero nominal interest rate into the indefinite future because inflation and output expectations (which lag behind actual inflation and output) would still be at levels that would imply an expectation that the ZLB should continue (indefinitely) to bind.

[44]In fact, one can show that all trajectories that begin in the region below both the ZLB line and the SM line in **Figure 1** converge eventually to the shaded region, where they remain forever, and diverge increasingly further from the deflationary steady state.

[45]Even if fiscal policy expectations are Ricardian, and the forecasts e_t are purely backward-looking as assumed above, an increase in government purchases should increase output and inflation, by increasing the term g_t in Equation 16. If a current increase in net government transfers does not reduce the present value of forecasted future net transfers—for example, because future primary surpluses are forecasted using an estimator such as that given in Equation 43—then an increase in net transfers will also increase output and inflation, by increasing the term b_{t+1} in Equation 18, while (if anything) also increasing the forecasts $\widehat{E}^i v^i_{t+1}$.

correct forecasting model may shift over time and then consequently place more weight on the most recent observations in their estimates of the current coefficients (see Sargent 1993 or Evans & Honkapohja 2001 for further discussion). In this case, the predicted TE dynamics are time invariant and can be characterized in terms of predicted unconditional moments (e.g., variances, autocovariances).

The dynamics of forecasts are again given by Equation 50 but now with the constant value γ substituted for γ_t; the implied TE dynamics of other variables are then given by Equation 48.[46] Equation 50 can alternatively be written in the form

$$e_t = \Lambda\, e_{t-1} + \lambda \xi_t, \tag{56}$$

where

$$\Lambda \equiv (1-\gamma)[I-\gamma A]^{-1}, \lambda \equiv \gamma[I-\gamma A]^{-1}a.$$

If policy satisfies Equation 24, both eigenvalues of A are inside the unit circle so that $[I-\gamma A]$ is invertible, Λ and λ are well defined, and both eigenvalues of Λ are also inside the unit circle. Hence Equation 56 defines stationary dynamics for the forecasts $\{e_t\}$ and consequently for the variables $\{\mathbf{x}_t\}$ as well.

Equation 50 implies that the forecasts will be serially correlated, even if the disturbances are i.i.d. More precisely, it implies that each element of e_t will be a linear combination of two first-order autoregressive processes (the innovations in which are generally correlated), with coefficients of serial correlation equal to the two eigenvalues of Λ. If γ is small (estimates are based on a fairly long history), the eigenvalues of Λ will be near 1, and these processes will be highly persistent. It then follows from Equation 48 that fluctuations in inflation and the output gap will have highly persistent components under the TE dynamics with learning. This contrasts sharply with the prediction of the REE analysis, according to which both inflation and the output gap should be serially uncorrelated if all fundamental disturbances are, as a consequence of Equation 25.

If the fundamental disturbances are instead themselves serially correlated, then persistent fluctuations in inflation and output are possible even under the REE dynamics. However, empirical NK models, such as those of Christiano et al. (2005) and Smets & Wouters (2007), generally find it necessary to introduce additional sources of persistence (indexation of prices and wages to past inflation, adjustment costs for expenditure), of debatable microeconomic realism, to fit the kind of persistence that is actually observed (see Woodford 2003, chapter 5, for discussion of the reasons for this). Learning dynamics provide an alternative potential source of intrinsic persistence, and some studies (e.g., Milani 2005, 2007, 2011; Slobodyan & Wouters 2009) find that there is less need for ad hoc structural persistence in econometric models that assume least-squares learning rather than RE.[47]

[46]The types of adaptive learning dynamics considered here are again of a fairly simple kind, as people's forecasting rules are assumed simply to forecast constant future values for each of the variables, and the only coefficients that must be learned are the estimated means of each variable. However, as Eusepi & Preston (2012b) discuss, even if one allows updating of the slope coefficients of a more complex linear regression model, in a local linear approximation to the implied TE dynamics, linearizing around the REE steady state, there are no additional dynamics resulting from the updating of the slope coefficients; the updating of the additional coefficients has only second-order effects on the TE dynamics.

[47]Eusepi & Preston (2011) similarly find that the introduction of learning dynamics results in a new channel for the propagation of the effects of technology shocks in an otherwise standard real business cycle model and argue that the model with learning produces fluctuations more similar to those in observed business cycles. Even larger departures from REE business cycle dynamics are predicted in the case of a model of learning that involves discrete switching between simple forecasting models of dramatically different character, as proposed by De Grauwe (2010).

4.3.2. Consequences for policy evaluation. The additional dynamics resulting from learning can change one's conclusions regarding the relative desirability of alternative monetary policy rules, even with respect to comparisons among rules that do not imply explosive learning dynamics. As a simple example, suppose that the central bank's short-run inflation target depends linearly on the cost-push shock,

$$\pi_t^* = \phi_u u_t, \tag{57}$$

for some $0 \leq \phi_\mu \leq 1$, and i_t is adjusted each period as necessary to ensure that $\pi_t = \pi_t^*$. The required interest rate can be determined, as a function of current disturbances and expectations, from Equations 16 and 17; these equations also indicate the implied evolution of the output gap.

Under the further assumption of RE beliefs, the model predicts that

$$y_t = -(1 - \phi_u)\kappa^{-1} u_t, \tag{58}$$

and hence that the unconditional variances of inflation and of the output gap will be

$$\mathrm{var}(\pi) = \phi_u^2 \sigma_u^2, \qquad \mathrm{var}(y) = (1 - \phi_u)^2 \kappa^{-2} \sigma_u^2,$$

where σ_u^2 is the variance of the cost-push shock. It follows that for all ϕ_u in this interval, increasing ϕ_u increases the volatility of equilibrium inflation but reduces the volatility of the equilibrium output gap. If policy is concerned with minimizing some weighted average of the two variances, the optimal choice of ϕ_u will be somewhere between the two extremes, at a point that depends on the relative weight on the two stabilization objectives.

If we instead assume adaptive learning of the kind specified by Equation 44, substitution of the policy rule in Equation 57 into the TE relation in Equation 17 implies that the output gap each period will be given by

$$y_t = -(1 - \phi_u)\kappa^{-1} u_t - (1 - \alpha)\beta\kappa^{-1} \widehat{p}_t^*, \tag{59}$$

where $\widehat{p}_t^*$ is the common forecast at date t of the value of p_{t+1}^*. The latter forecast will be given by Equation 46; if the estimates of the means of each of the variables evolve in accordance with Equation 44, for some $0 < \gamma < 1$ this implies that

$$\widehat{p}_t^* = (1 - \gamma)\widehat{p}_{t-1}^* + \gamma\left[(1 - \alpha\beta)^{-1} \pi_t + \xi y_t + \mu_t\right].$$

Substituting Equations 57 and 59 for π_t and y_t, respectively, in this expression yields a law of motion for the forecast of the form

$$\widehat{p}_t^* = \rho \widehat{p}_{t+1}^* + \phi_u \psi u_t, \tag{60}$$

where both

$$0 < \rho \equiv \frac{(1 - \gamma)(1 - \alpha\beta)}{1 - (1 - \gamma)\alpha\beta} < 1 \quad \text{and} \quad \psi \equiv \frac{\gamma}{\alpha[1 - (1 - \gamma)\alpha\beta]} > 0$$

are independent of the choice of ϕ_u.

Equation 60 implies that if $\phi_u > 0$, a positive cost-push shock immediately raises the forecast $\widehat{p}_t^*$, and the forecast continues to be higher in subsequent periods as well (to an extent that decreases exponentially over time). Comparing Equation 59 with the REE prediction in Equation 58, we see that with adaptive learning, the output reduction in the period of the shock is greater than would occur under RE; moreover, the negative effect on the output gap persists, rather than being limited to the period of the shock. For both reasons, a given value of $\phi_u > 0$ does not reduce the predicted variance of the output gap as much as is predicted by the REE analysis, although it continues to increase the predicted variance of inflation by the same amount. Thus the trade-off between inflation stabilization and output-gap stabilization is steeper in the case of learning: Less reduction in the variance of the output gap is achieved by a given increase in the variance of inflation. The implication, as argued by Orphanides & Williams (2005), is that under given preferences with regard to inflation and output-gap stability, it will be optimal to choose a lower value of ϕ_u (maintaining tighter control of inflation) when one recognizes that people must learn to forecast macroeconomic conditions, relative to what one would conclude from the REE analysis.[48]

5. TEMPORARY EQUILIBRIUM DYNAMICS WITH NEARLY CORRECT BELIEFS

There is, however, another way of ensuring that one's model's predictions do not depend on a supposition that people will fail to notice patterns in the data that should actually be easily discerned. These alternative approaches are based not on an explicit specification of the procedure used to look for such patterns, as in the case of econometric learning models, but rather on a direct requirement that probability beliefs, however obtained, not be too different from the true probabilities (according to one's model). Approaches of this kind propose no model of how people reach the probability beliefs they hold, but instead focus on defining the respects in which subjective beliefs should reasonably be expected to be similar to objective probabilities, and the other respects in which one might expect more variation in subjective beliefs. In this section, I discuss two examples of how this might be done: the rational belief equilibria of Kurz and coauthors (Kurz 1994, 1997, 2012; Kurz & Motolese 2011) and the near-rational expectations proposed by Woodford (2010) and explored further by Adam & Woodford (2012).

Before describing these approaches, it is important to note that a hypothesis that beliefs are nearly correct does not imply that they are nearly the same as (any possible) REE beliefs. The extent to which beliefs are correct depends on their conformity with the actual TE dynamics, which may differ greatly from REE dynamics, and not on their conformity with REE predictions. This difference is emphasized in particular by Kurz (2012), who demonstrates the possibility of sizeable aggregate fluctuations even when the magnitude of exogenous disturbances to fundamentals is much smaller than must be postulated to account for the fluctuations using DSGE models that assume RE.

5.1. Rational Belief Equilibria

Kurz (1994) proposes a relaxation of the RE hypothesis in which the probability beliefs of decision makers are required to imply model-consistent values for some data moments, but not for all the data moments that are relevant to their forecasts and hence to their decisions. Certain quantities (including conventional macroeconomic aggregates, such as the rate of growth of GDP or the Consumer Price Index) are assumed to be objectively measurable, and as a consequence everyone is

[48]Orphanides & Williams (2005) consider a one-parameter family of policies similar to the one considered here but in the context of a simpler model of the way in which expectations affect aggregate supply. For implications of learning dynamics for the optimal choice of a policy rule within more complex families of candidate policies, readers are referred to Gaspar et al. (2011).

assumed to agree about the current and past values of these variables. The postulate of rational beliefs (RB) then requires that in any stationary equilibrium [a rational belief equilibrium (RBE)] consistent with some time-invariant policy, everyone must also agree about all the unconditional first and second moments[49] of these objectively measurable variables and assign values to these moments that coincide with the predictions of the model about this particular RBE.[50]

But these variables are not the only ones on the basis of which individuals form their forecasts; there are also subjective variables (belief states) about which they need not agree. A given decision maker is assumed to have coherent probability beliefs about the joint distribution of his or her own belief states and the objectively measurable variables, on the basis of which the belief states modify his or her forecasts of the future paths of the objectively measurable variables, but these data moments need not be ones about which others agree, and the probability beliefs of an individual need not coincide in this respect with the predictions of the model. It is in this latter respect that the RB postulate is weaker than RE. Insofar as people are assumed to learn the correct values of some data moments but not others, the RBE concept is a cousin of the RPE concept discussed in Section 4.1.

The content of the RB postulate, as well as the sense in which it is weaker than RE, is best illustrated using an example. Suppose that the natural rate of output is the sum of two components,

$$Y_t^n = \overline{Y}_t^n + \xi_{2t}, \tag{61}$$

where the permanent component $\overline{Y}_t^n$ evolves as a random walk,

$$\overline{Y}_t^n = \overline{Y}_{t-1}^n + \xi_{1t}, \tag{62}$$

with $\{\xi_{1t}\}$ an i.i.d. innovation distributed as $N(0, \sigma_1^2)$, and the transitory component ξ_{2t} another i.i.d. innovation, distributed as $N(0, \sigma_2^2)$ and independent of the permanent shocks. If the process $\{Y_t^n\}$ is objectively measurable but its permanent and transitory components are not, and no other objectively measurable variables provide information about this decomposition, then an optimal estimate of the permanent component (or optimal forecast of the long-run level) of Y_t^n at any time t is given by an exponentially weighted moving average[51]

$$\overline{Y}_t \equiv (1-\lambda) \sum_{j=0}^{\infty} \lambda^j Y_{t-j}^n,$$

where the smoothing factor $0 < \lambda < 1$ is given by

$$\lambda = \frac{2}{2 + q + \sqrt{q^2 + 4q}}, \quad q \equiv \frac{\sigma_1^2}{\sigma_2^2}.$$

Conditional only on objectively measurable data, then an optimal forecast of the future natural rate at any horizon $k \geq 0$ will be given by

[49]By "all second moments," I mean to include all covariances between leads and lags of the various variables.

[50]In fact, Kurz (1994) proposes the stronger postulate that the subjective assessment of the unconditional joint distribution of the objectively measurable variables must coincide with their model-implied unconditional distribution. In the case of a linear model with additive Gaussian disturbances, of the kind used in applications such as Kurz (2012), and in the example presented below, the identity of the two unconditional distributions is equivalent to the identity of the complete list of first and second moments.

[51]This corresponds to the Bayesian posterior mean, or minimum-mean-squared-error forecast, using a Kalman filter (Harvey 1989, chapter 4), as originally derived by Muth (1960).

$$\overline{E}_t Y^n_{t+k} = \overline{Y}_t. \tag{63}$$

Suppose, however, that in addition to the objectively measurable data, each individual price setter j has a subjective estimate of the permanent component, which I denote z^j_t. If each price setter correctly understands the laws of motion given in Equations 61 and 62, this implies that subjective forecasts will be given by

$$\widehat{E}^j_t Y^n_{t+k} = z^j_t \tag{64}$$

rather than by Equation 63. Note that each individual's beliefs are described by a completely specified, internally consistent probability measure, which is moreover consistent with the true first and second moments of all objectively measurable data; for example, these beliefs imply correct (model-consistent) values for the unconditional moments $\mathrm{E}[\Delta Y^n_t]$ and $\mathrm{cov}(\Delta Y^n_t, \Delta Y^n_{t-k})$ for all k, as these can be derived from Equations 61 and 62.[52]

But individual beliefs about the statistics of the subjective belief state z^j_t and its comovement with objectively measurable data need not coincide with the beliefs of others, or with the way in which the model describes the evolution of these variables. In our example, although each individual j believes that $z^j_t = \overline{Y}^n_t$, it does not follow that z^j_t must take the same numerical value for all j. Moreover, even if, as in Kurz & Motolese (2011) and Kurz (2012), we suppose that the population distribution of subjective beliefs and hence the population mean $Z_t \equiv \int z^j_t dj$ are objectively measurable data, it does not follow that z^j_t must equal Z_t for each individual. Individuals can be aware that their personal estimate z^j_t differs from the average estimate, without any internal inconsistency of their beliefs. The RB postulate requires that all people have model-consistent beliefs about unconditional moments such as $\mathrm{E}[Y^n_t - Z_t]$, $\mathrm{E}[\Delta Z_t]$, $\mathrm{cov}(\Delta Y^n_{t+k}, Y^n_t - Z_t)$, and so on. But awareness of these moments and observation of Z_t do not give price setters any reason to doubt the validity of their forecasts in Equation 64, given their beliefs in the laws of motion in Equations 61 and 62 and their beliefs that z^j_t is an accurate (but personal) observation of the value of $\overline{Y}^n_t$. (This simply requires each individual to believe that others' personal assessments of the value of the permanent component are erroneous, even though he or she understands that they each believe their personal assessments to be correct.[53])

As one possible example of how this makes possible an additional source of aggregate fluctuations, suppose that people's subjective assessments are given by $z^j_t = \overline{Y}_t + \nu^j_t$, where ν^j_t is a random term that evolves independently of all fundamental variables, including both the permanent and transitory components of Y^n_t. Thus, because $\overline{Y}_t$ is objectively measurable, the subjective state z^j_t reflects no additional information about the future evolution of the natural rate (or any other fundamentals). Moreover, suppose that the errors ν^j_t are correlated across individuals so that the aggregate error $\nu_t \equiv \int \nu^j_t dj$ is not equal to 0. Then because ν_t represents an error in the average estimate of a variable that is relevant for pricing decisions (the error in the average estimate

[52]Kurz (1994) refers to these common beliefs not as correct beliefs about unconditional moments, but only as "empirical frequencies." However, in applications such as Kurz & Motolese (2011) and Kurz (2012), the calculations used to explain or predict data are carried out under the assumption that the empirical frequencies correspond to model-implied unconditional moments, under time-invariant stochastic processes for the various disturbances specified in the model.

[53]Thus this equilibrium concept allows a much wider range of possible specifications of belief dynamics than an RE model with private information, of the kind considered by Rondina & Walker (2012). Rondina & Walker assume that all individuals agree about the joint distribution of all publicly or privately observable variables, although individuals do not observe other individuals' private signals about the separate components of the aggregate disturbance. In Kurz's work, instead, people agree to disagree. It is therefore not necessary to suppose that there is anything secret about individuals' subjective beliefs; it is only the basis for accepting these subjective assessments as correct that is not shared.

$\int \widehat{E}_t^j\left[\overline{Y}_t^n - \overline{Y}_t\right]$), it will affect the determination of endogenous aggregate variables, such as output and inflation, and variation in ν_t will be an additional source of variability in these variables, in addition to the random variation in fundamentals such as $\{\xi_{1t}, \xi_{2t}\}$.

To illustrate the effects of fluctuations in the aggregate belief state on endogenous variables, let the monetary policy rule be specified by a target criterion: That is, the central bank adjusts its instrument as necessary to ensure that the linear relationship

$$\pi_t + \phi\left(Y_t - \overline{Y}_t\right) = 0 \tag{65}$$

holds at all times.[54] This represents a form of flexible inflation targeting (Svensson 1999), in which the concept of the output gap in the central bank's target criterion is output relative to the central bank's estimate of long-run potential, rather than relative to the current natural rate of output. As a simple example in which belief fluctuations provide an independent source of aggregate variability, suppose that ν_t evolves as an AR(1) process,

$$\nu_t = \rho\nu_{t-1} + \epsilon_t, \tag{66}$$

where $0 \geq \rho < 1$ and $\{\epsilon_t\}$ is an i.i.d. innovation, with distribution $N(0, \sigma_\epsilon^2)$ independent of all fundamental states. We can then solve for the equilibrium dynamics of inflation and output implied by the TE relations, Equation 65, and the above assumptions about subjective expectations, using the method of undetermined coefficients.

Let us conjecture beliefs on the part of each price setter j of the form

$$\widehat{E}_t^j p_{t+1}^{*j} = \gamma_1\left(\overline{Y}_t - z_t^j\right) + \gamma_2\left(\overline{Y}_t - Z_t\right), \tag{67}$$

for coefficients γ_1 and γ_2 that remain to be determined. Average beliefs are then given by

$$\int \widehat{E}_t^j p_{t+1}^{*j} dj = \gamma\left(\overline{Y} - Z_t\right),$$

where $\gamma \equiv \gamma_1 + \gamma_2$. Let us also suppose for simplicity that the cost-push shock u_t is equal to 0 at all times.[55] The TE relation in Equation 17 then implies that inflation and the output gap must be given by

$$\pi_t = \frac{\kappa\phi}{\kappa+\phi}\left(\overline{Y}_t - Y_t^n\right) - \frac{(1-\alpha)\beta\gamma\phi}{\kappa+\phi}\nu_t, \tag{68}$$

$$y_t = \frac{\phi}{\kappa+\phi}\left(\overline{Y}_t - Y_t^n\right) + \frac{(1-\alpha)\beta\gamma}{\kappa+\phi}\nu_t. \tag{69}$$

Because Equations 68 and 69 are relationships among objectively measurable variables,[56] the RB postulate requires that the subjective probability beliefs of each price setter be consistent with

[54]The state-contingent path for the interest rate i_t required for Equation 65 to hold each period will depend on the subjective expectations of both price setters and consumers. I suppose here that the central bank observes average expectations when setting i_t and so can implement a reaction function that makes Equation 65 a necessary consequence of the TE relations that determine π_t and y_t, regardless of what those subjective expectations may be. Thus Equation 65 can be treated as an equilibrium relation in solving for the equilibrium dynamics under a given hypothesis about expectations.

[55]Note that even under this assumption, the model implies the existence of equilibrium fluctuations in inflation and in the output gap, owing to the discrepancy between the concept of potential output $\left(\overline{Y}_t\right)$ used in the central bank's target criterion (Equation 65) and the one (Y_t^n) that shifts the AS relation (Equation 17).

[56]Recall that $\nu_t = Z_t - \overline{Y}_t$, and the average belief state Z_t is assumed to be objectively measurable.

them. These relations, together with the laws of motion in Equations 61, 62, and 66 for the exogenous aggregate state variables, further imply that correct forecasts of future inflation and output are given by

$$E_t \pi_{t+k} = \frac{\kappa\phi}{\kappa+\phi}\lambda^k\left(\overline{Y}_t - \overline{Y}_t^n\right) - \frac{(1-\alpha)\beta\gamma\phi}{\kappa+\phi}\rho^k \nu_t,$$

$$E_t y_{t+k} = \frac{\phi}{\kappa+\phi}\lambda^k\left(\overline{Y}_t - \overline{Y}_t^n\right) + \frac{(1-\alpha)\beta\gamma}{\kappa+\phi}\rho^k \nu_t,$$

for any horizon $k \geq 1$, where $E_t[\cdot]$ refers to the expectation conditional on the history of all exogenous states up through period t, including the (unobserved) value of $\overline{Y}_t^n$. If price setters are assumed to correctly understand these laws of motion,[57] then their subjective forecasts of future inflation and output gaps must conform to these equations as well, but with the value of $\overline{Y}_t^n$ replaced by each individual's subjective estimate of this state. Thus for any price setter j, each of the forecasts is a linear function of $\overline{Y}_t - z_t^j$ and $\nu_t \equiv Z_t - \overline{Y}_t$.[58]

Substituting these subjective forecasts into the definition given in Equation 12, we can obtain an expression for $\widehat{E}_t^j p_{t+1}^{*j}$ as a linear function of $\overline{Y}_t - z_t^j$ and $\overline{Y}_t - Z_t$, as conjectured in Equation 67. We now, however, have expressions for the coefficients γ_1 and γ_2 (given in the **Supplemental Appendix**) as functions of the assumed value of γ. Requiring the implied values of these coefficients to equal their conjectured values, one yields two linear equations to solve for the unknown coefficients γ_1 and γ_2. As shown in the **Supplemental Appendix**, our sign assumptions on parameters imply the existence of a unique solution, with $\gamma_1 > 0$, $\gamma_2 \geq 0$, and $\gamma_2 > 0$ if $\rho > 0$.

Supplemental Material

We thus obtain TE dynamics consistent with the RB postulate in which fluctuations in the aggregate belief state ν_t cause random variations in inflation and output. It is instructive to compare this solution with the REE dynamics under the policy rule given in Equation 65. We may assume as above that each individual observes a personal state variable z_t^j (a gut feeling, if one likes) that is distributed as assumed above, but under the RE hypothesis, each individual must correctly understand the joint distribution of z_t^j and all other variables. This would mean correctly understanding that z_t^j contains no information that is useful for predicting the future path of the natural rate of output (given that $\overline{Y}_t$ is independently observable) and similarly that Z_t is uninformative. RE forecasts of all variables would then correspond simply to the expectations of those variables conditional on the observed history of the natural rate of output; thus, for example, the common forecast of the future natural rate of output would be given by Equation 63. It is shown in the **Supplemental Appendix** that under the policy rule in Equation 65, the unique stationary REE is one in which inflation and the output gap are given by Equations 68 and 69 but with $\gamma = 0$, whereas $\gamma > 0$ in the RBE discussed above. Thus the RBE beliefs do not change the

[57]The RB postulate requires that all price setters correctly understand the autocorrelation function of the objectively measurable process $\{\nu_t\}$, but it does not require that they agree that an unbiased forecast of ν_{t+k} at time t depends only on the current value ν_t; they may have subjective judgments about the likely future path of the aggregate belief state that they believe are more accurate than the forecast that could be made on the basis of objectively measurable data alone. Here I make the more restrictive assumption that no one believes they have additional insight into the future evolution of any exogenous states except for believing in their personal estimates of the permanent component $\overline{Y}_t^n$.

[58]Kurz & Motolese (2011) state that "those who believe the economy is stationary" will necessarily forecast using the "empirical measure"—that is, using only the information contained in the history of objectively measurable variables, and so make forecasts such as Equation 63. But in fact, the subjective probability beliefs specified here imply that $\left\{\pi_t, y_t, \overline{Y}_t - \overline{Y}_t^n, Y_t^n - \overline{Y}_t^n, \overline{Y}_t - Z_t, \Delta\overline{Y}_t^n\right\}$ are jointly stationary processes, and the same is true of the RBE beliefs specified in the applications proposed by Kurz & Motolese (2011) and Kurz (2012). The crucial issue is actually not stationarity, but whether variables other than objectively measurable ones are also used in forecasting.

response of inflation or output to exogenous fluctuations in the natural rate of output but result in increased variability of both inflation and the output gap for any value of ϕ, relative to the REE prediction, owing to the existence of fluctuations unrelated to any changes in fundamentals, purely because of variation in the aggregate belief state.

With the above simple calculation, it may appear that the RBE hypothesis makes definite quantitative predictions about the evolution of endogenous variables under a given policy rule, but this is actually not true; the RBE constructed above is only one possible example of TE dynamics consistent with the RB postulate under the assumed policy rule. First of all, there is an RBE of the kind assumed above for any specification of the serial correlation coefficient ρ and of the innovation variance σ_{ϵ}^2 for the process $\{v_t\}$. Moreover, there is no reason why $\{v_t\}$ must be an AR(1) process; this allowed us to verify the conjecture that subjective forecasts were of the form of Equation 67, but we might equally well have assumed a more complex process for $\{v_t\}$ and still solved for an RBE, in which, however, subjective forecasts would be correspondingly more complex. Thus if we allow $\{v_t\}$ to be any process in some larger parametric family, we can obtain a multiparameter family of RBE associated with the given policy in Equation 65. But even this understates the multiplicity of possible RBE. For it was not necessary to have assumed that people believe that they have an additional (personal) awareness of the decomposition of Y_t^n into permanent and transitory components but no additional personal insight into the economy's future evolution of any other sort. Allowing for other types of subjective beliefs (which need not be correlated with actual outcomes, according to one's model, in the way that people believe they are) would further expand the set of RBE solutions consistent with a given policy rule.

Kurz and coauthors argue that the more flexible relationship between the evolution of exogenous fundamentals and that of endogenous variables allowed by this relaxation of the RE hypothesis can make sense of some of the empirical difficulties faced by RE models. For example, Kurz & Motolese (2011) discuss RBE of an asset-pricing model in which there is a risky asset in fixed supply and an exogenously given riskless rate of return (independent of the quantity invested in the riskless asset). The dividend on the risky asset is an exogenous process, about the future evolution of which individual investors believe they have additional personal insight, beyond the information contained in the past history of the dividend, just as in the case of subjective forecasts of the natural rate of output in the above example. In the RBE, variations in the aggregate belief state become an additional source of variation in equilibrium asset prices and in particular result in a time-varying risk premium of the kind that is found to be empirically important in many asset markets.

Kurz & Motolese (2011) estimate the parameters of their model using data on term premia associated with federal funds futures and Treasury bills and find that allowance for the more flexible class of equilibria allows the data to be fit better; their best-fitting RBE implies that more than half of the measured risk premia result from fluctuations in the aggregate belief state. This suggests that the kind of additional flexibility allowed by the concept of an RBE may be of empirical relevance. At the same time, because the predictions of the more general theory are much less specific, it is not obvious that findings such as those of Kurz & Motolese can be regarded as confirming a specific theoretical view of the nature of subjective beliefs.[59]

[59]Although the sets of possible RBE discussed in papers such as Kurz & Motolese (2011) and Kurz (2012) involve only a few free parameters, this is not because the RB postulate alone allows one to derive such specific conclusions—a large number of additional (theoretically unmotivated) assumptions are made as well to obtain equilibria of a particular form. Moreover, in neither of these applied papers are all the restrictions implied by the RB postulate imposed; the solutions proposed as possible accounts of actual data are actually examples of an even weaker version of the RBE concept, and the proposed restrictions on the stochastic processes characterizing the data are mainly adopted for convenience rather than following from a conception of rationality.

Relaxation of the RE hypothesis also has potential consequences for policy design; as illustrated by the above example, the degree of macroeconomic stability guaranteed by commitment to a given policy rule may not be as great as a mere analysis of the REE dynamics consistent with it would suggest. This raises the possibility that alternative rules might provide more robust approaches to stabilization, even if they do not lead to a superior REE. Although there will not be a unique RBE consistent with a given policy rule, or even a unique RBE associated with a given restricted state space (as in the analysis of the minimum-state-variable REE above), it may be possible to compare the data moments associated with the entire range of possible RBE for alternative parameterizations of a policy rule. Kurz (2012) undertakes an illustrative analysis of this kind of the consequences of alternative central-bank reaction functions in the context of an NK model closely related to the one presented here. However, the comparisons undertaken consider only certain parametric classes of RBE, and it is unclear why attention should be restricted to these specific types of equilibria. This seems an important limitation of the Kurz approach for purposes of policy analysis. The alternative approach presented next instead allows a clear delineation of the set of equilibria consistent with a given policy rule.

5.2. Near-Rational Expectations

Rather than distinguishing a priori between data moments that individuals should correctly assess and those that they may not, depending on the nature of the variable in question, the assumption of near-rational expectations (NRE) in Woodford (2010) instead defines a set of probability beliefs that are close enough to the predictions of one's model to be plausibly held by decision makers in such a situation, on purely statistical grounds. Essentially, an alternative probability distribution is close to the predicted probabilities of outcomes in a given equilibrium if the alternative distribution represents a sample distribution of outcomes that could be observed in some finite number of repetitions of the equilibrium. This requires, for example, that near-rational subjective expectations assign zero probability to all outcomes (definable in terms of state variables in only a finite number of periods) that occur with zero probability in equilibrium.

This means that each agent's subjective probability measure over possible paths for all variables must be absolutely continuous (over finite intervals) with respect to the equilibrium probability measure. This in turn implies there must exist a scalar stochastic process $\{m_t\}$ for each agent—the agent's belief distortion factor—with $m_t \geq 0$, $E_t m_{t+1} = 1$, at all times such that the agent's subjective one-period-ahead forecast of any variable X_{t+1} is given by

$$\widehat{E}_t X_{t+1} = E_t\left[m_{t+1} X_{t+1}\right], \tag{70}$$

where $E_t[\cdot]$ indicates the conditional expectation under the true (model-implied) probabilities in the particular equilibrium. Thus a value $m_{t+1} > 1$ in a particular state of the world at date $t+1$ implies that, conditional on reaching the predecessor state at date t, the agent exaggerates the probability of reaching this state relative to the correct equilibrium probability. Internal consistency of individual probability beliefs then implies that longer-horizon subjective forecasts are correspondingly given by

$$\widehat{E}_t X_{t+j} = E_t\left[m_{t+1} \cdots m_{t+j} X_{t+j}\right].$$

The degree of discrepancy between subjective and objective probability beliefs can then be measured by the degree to which the distortion factor $\{m_t\}$ differs from a constant factor, equal to 1 in all states. A measure of the degree of discrepancy in one-period-ahead beliefs (looking forward from any period t) with appealing properties is the relative entropy between the subjective and objective conditional probabilities

$$R_t \equiv E_t\left[m_{t+1} \log m_{t+1}\right]. \tag{71}$$

This is the nonnegative convex function of the belief distortion factor that achieves its minimum possible value of 0 if and only if $m_{t+1} = 1$ almost surely (the RE case). Moreover, the probability of observing a sample frequency distribution for the possible outcomes at date $t+1$ that is close to any given subjective measure, in the case of a large (but finite) number of independent draws from the equilibrium probability measure, is (in the case of a large-enough number of draws) a decreasing function of the relative entropy of the subjective measure (see, e.g., Cover & Thomas 2006). Hence subjective beliefs under which R_t is small (although positive) each period are ones that could plausibly be maintained even by an agent with considerable experience of typical equilibrium outcomes.

Woodford (2010) accordingly defines an equilibrium with NRE as a situation in which each agent optimizes on the basis of internally consistent probability beliefs for which R_t is sufficiently small each period, when calculated with respect to the equilibrium probability measure describing the outcomes resulting (in each possible state of the world) from their collective choices. Note that the equilibrium measure will not generally be the REE measure because people act on the basis of non-REE beliefs; hence near-rational expectations equilibrium (NREE) outcomes need not be near the REE outcomes for beliefs to be near-rational.

In the context of the NK model described above, an NREE corresponds to stochastic processes $\{\bar{v}_t^i\}$ and distortion factors $\{m_t^i\}$ for each household, and processes $\{p_t^{*j}\}$ and distortion factors $\{m_t^j\}$ for each firm, such that Equations 10 and 15 hold each period when subjective forecasts are given by Equation 70 for each agent, and the distortion factors imply that the relative entropy for each agent remains within some bound. If, for example, we assume a monetary policy rule of the form of Equation 21 and restrict attention to the special case of common subjective probability beliefs for all agents, then an NREE corresponds to a vector stochastic process $\{z_t\}$ and distortion factor $\{m_t\}$ such that

$$z_t = B\, E_t\left[m_{t+1} z_{t+1}\right] + b\xi_t \tag{72}$$

holds each period (where B and b are again the matrices in Equation 22), and the relative entropy (Equation 71) implied by the distortion factor satisfies the specified bound.

5.3. Robustly Optimal Policy

For any positive upper bound on the allowable relative entropy, the set of NREE consistent with a given policy rule will be large. How then can this kind of theory provide a basis for selection of a particular policy rule? Woodford (2010) proposes that one choose the policy that implies the highest possible lower bound for one's welfare objective (or lowest possible upper bound for one's loss function), across the entire set of NREE consistent with the rule, under some specified bound on the allowable size of belief distortions. Such a maximin approach to policy choice is in the spirit of the robust control approach to dealing with model uncertainty advocated by Hansen & Sargent (2008).

This approach requires one to determine, for any candidate policy rule, the worst-case belief distortion process, which implies an NREE that is the worst possible for the policy maker's welfare objective, subject to the bound on the size of belief distortions that are contemplated. As an example, suppose that the objective of policy is to minimize a discounted loss function of the form

$$E_0 \sum_{t=0}^{\infty} \beta^t \left[\pi_t^2 + \lambda (y_t - y^*)^2 \right] \tag{73}$$

for some relative weight $\lambda > 0$ on output-gap stabilization, and some optimal output gap $y^* > 0$, as in Clarida et al. (1999). Here the expectation $E_0[\cdot]$ used to define the objective refers to the probability beliefs of the policy maker, which need not be shared by others. And let us again consider policy commitments in the simple family given in Equation 57.

If for simplicity we restrict attention to equilibria in which belief distortions are common to all agents, Equations 11 and 20 imply that

$$\pi_t = \kappa y_t + u_t + \beta E_t \left[m_{t+1} \pi_{t+1} \right]$$

each period, which is just the NRE generalization of the NK Phillips curve assumed by Clarida et al (1999). Substituting Equation 57 for the path of inflation, this implies that the output gap must satisfy

$$y_t = -\kappa^{-1} \left\{ (1 - \phi_u) u_t + \beta \phi_u E_t \left[m_{t+1} u_{t+1} \right] \right\} \tag{74}$$

in the NREE corresponding to any distortion process $\{m_t\}$.

Because the path of inflation is independent of belief distortions under a policy commitment of the hypothesized type, the belief distortions that maximize Equation 73 involve a choice of the one-period-ahead distortion factors $\{m_{t+1}\}$ looking forward from any date t so as to maximize $(y_t - y^*)^2$ subject to an upper bound

$$R_t \leq \overline{R}, \tag{75}$$

where R_t is defined in Equation 71, and the size of $\overline{R} > 0$ indicates the allowable degree of departure from model consistency. Hence the factors $\{m_{t+1}\}$ should be chosen to maximize

$$\frac{1}{2}(y_t - y^*)^2 + \theta_t E_t \left[m_{t+1} \log m_{t+1} \right]$$

subject to the constraint that $E_t m_{t+1} = 1$, where y_t is given by Equation 74, and θ_t is a Lagrange multiplier associated with the constraint in Equation 75.

If u_{t+1} is i.i.d. $\mathcal{N}(0, \sigma_u^2)$, the solution to this problem is easily shown to involve a state-contingent distortion factor

$$\log\ m_{t+1} = \alpha + \gamma u_{t+1},$$

where

$$\alpha = -\overline{R}, \quad \gamma = \pm \frac{\left(2\overline{R}\right)^{1/2}}{\sigma_u}.$$

The positive root for γ is optimal if $y_t < y^*$ (the most common case), whereas the negative root is optimal if $y_t > y^*$. (When $y_t < y^*$, the policy maker's trade-off is made even more painful by an increase in inflation expectations, which shift the short-run Phillips curve in a way that increases the tension between the goals of keeping inflation near 0 and the output gap near y^*, and expected inflation is increased if people exaggerate the likelihood of positive cost-push shocks. If $y_t > y^*$, instead the worst-case belief distortions would be ones that reduce inflation expectations, by exaggerating the likelihood of negative cost-push shocks.)

The worst-case beliefs then imply

$$\mathrm{E}_t[m_{t+1}u_{t+1}] = \gamma\sigma_u^2 = \pm(2\overline{R})^{1/2}\sigma_u,$$

taking care to select the root that implies the largest gap between y_t and y^*. It follows that

$$|y_t - y^*| = |y^* + (1-\phi_u)\kappa^{-1}u_t| + \phi_u\beta\kappa^{-1}(2\overline{R})^{1/2}\sigma_u \quad (76)$$

for all realizations of u_t. Equation 76 shows that increasing ϕ_u reduces the sensitivity of $y_t - y^*$ to cost-push shocks (as in the RE analysis) but at the cost of making it possible for the absolute value of the gap in the absence of any cost-push shock to be larger as a result of belief distortions.

The upper bound for Equation 73 in the case of any policy in the simple family (Equation 57) is then given by

$$(1-\beta)^{-1}\left[L_\pi + \lambda L_y^{pess}\right],$$

where

$$L_\pi \equiv \mathrm{E}\left[\pi_t^2\right] = \phi_u^2\sigma_u^2$$

is the same function of ϕ_u as in the RE analysis, and

$$L_y^{pess} \equiv \mathrm{E}\left[(y_t - y^*)^2\right]$$

is evaluated under the worst-case belief distortions; in both expressions, the expectation is over possible realizations of u_t. One observes that $\partial L_\pi/\partial\phi_u = 2\phi_u\sigma_u^2$, regardless of the degree of concern for robustness, but one finds that allowance for belief distortions ($\overline{R} > 0$) makes the value of the derivative $\partial L_y^{pess}/\partial\phi_u$ less negative (or more positive) at each value of ϕ_u. Hence the value of ϕ_u at which the marginal reduction in expected losses with respect to output-gap stabilization no longer outweighs the marginal increase in expected losses with respect to inflation stabilization will be reached at a lower value of ϕ_u under the worst-case beliefs (when $\overline{R} > 0$) than under the RE analysis. In fact, the upper bound for expected losses may be minimized at $\phi_u = 0$, whereas complete inflation stabilization is never optimal under the RE analysis, and the optimal ϕ_u remains bounded away from 1, regardless of how large the weight λ on the output-gap stabilization objective may be, whereas the optimal ϕ_u approaches 1 as $\lambda \to \infty$ in the RE analysis.

The robustly optimal policy within this simple family thus involves greater stability of inflation in the face of cost-push shocks than would be optimal if one could be sure that people would have model-consistent expectations. This is similar to the conclusion obtained in Section 4.3.2 when analyzing alternative policies under the assumption of adaptive learning, and again the basic reason is that variations of inflation in response to cost-push shocks make it too easy for people to misestimate the average future rate of inflation, causing undesirable instability in the short-run Phillips curve trade-off.

The reasons for inflation expectations to be insufficiently well anchored are somewhat different in the two cases: In the learning analysis, it was assumed that inflation expectations necessarily drift in response to certain observations of inflation outcomes, and a large value of ϕ_u was dangerous because it increased the frequency of occurrence of observations that would lead to significant expectational errors; here, instead, no precise prediction is made about what expectations must be, but a large value of ϕ_u is dangerous because it allows more significant expectational errors to be consistent with the assumed bound on relative entropy. Yet ultimately, the problematic feature of the large-ϕ_u policy is the same in both cases: It makes sample paths in which average observed inflation differs significantly from the actual long-run inflation target (in particular, paths in which the sample average is significantly higher) occur too frequently.

Woodford (2010) extends the above analysis by considering a much more flexible family of policies, in which the short-term inflation target π_t^* is an arbitrary linear function of the history of cost-push shocks, and shows how the optimal commitment of this form differs from the optimal commitment in the RE analysis of Clarida et al. (1999). As in the simpler exercise above, the robustly optimal policy commitment involves a lower amplitude of inflation surprises in response to cost-push shocks; Woodford shows that it also involves a greater degree of commitment to the subsequent reversal of any effects on the price level of past cost-push shocks.[60] Kwon & Miao (2012) show how a similar method can be used to characterize the robustly optimal policy commitment for a broad class of linear-quadratic policy problems and generalize Woodford's (2010) results to the cases of persistent cost-push shocks and of a more general form of aggregate-supply relation that incorporates intrinsic inflation inertia.

Adam & Woodford (2012) further extend the analysis of Woodford (2010), considering policy commitments that are not necessarily expressed in terms of inflation targets that depend only on the history of exogenous disturbances. They find that the conclusions mentioned above continue to hold, as a description of how inflation must be expected to evolve in response to cost-push shocks under the worst-case beliefs, even if a robustly optimal policy commitment (within the more general family) need not require inflation to evolve this way regardless of the nature of belief distortions. In Adam & Woodford's analysis, there is not a uniquely defined policy rule that is robustly optimal; instead, there exists a large class of policy rules that imply the same dynamics under the worst-case belief distortions and hence achieve the same minimum upper bound for the loss function, although they may be associated with different TE dynamics under other kinds of distorted beliefs that are also consistent with the relative-entropy bound.[61]

Among the robustly optimal policy rules is one that involves commitment to a target criterion: The central bank uses its policy instrument to ensure that the joint evolution of inflation and output satisfies a linear relationship of the form

$$\pi_t + \phi_s(\pi_t - E_{t-1}\pi_t) + \phi_y(y_t - y_{t-1}) = 0 \tag{77}$$

each period, where ϕ_s and ϕ_y are both positive coefficients, which depend both on model parameters and on the relative weight λ assumed in the objective given in Equation 73,[62] and $E_{t-1}\pi_t$ indicates the policy maker's forecast of inflation a period earlier. Here the coefficient $\phi_s >$ 0 multiplying the inflation surprise results from the concern for robustness, and this coefficient is larger when the concern for robustness (as measured by the relative-entropy bound) is greater. The

[60]This kind of robust policy problem is compared to alternative ways of introducing robustness to uncertainty about the correctness of model equations into an optimal monetary stabilization policy problem, in the context of the same linear-quadratic NK framework used here, in Hansen & Sargent (2012).

[61]It should be recalled that also under the RE analysis, the optimal policy commitment is not uniquely defined. Instead, the optimal REE dynamics are uniquely defined, while there are many different policy rules that can achieve these dynamics as a determinate equilibrium outcome; the rules differ in the behavior that they prescribe out of equilibrium, although the policy instrument evolves in the same way in equilibrium under each of them. Under the robust policy analysis, the different robustly optimal rules also differ in the sets of possible equilibrium outcomes associated with them because a given rule does not imply a determinate equilibrium, except under a particular specification of the belief distortions.

[62]Adam & Woodford (2012) also show how to characterize welfare-maximizing policy, when welfare is defined by the expected utility (under the policy maker's expectations) of a representative household. There is again a robustly optimal target criterion of the form of Equation 77, the coefficients of which now depend purely on model parameters.

presence of this term reduces the extent to which a cost-push shock should be allowed to cause a surprise change in the rate of inflation, as it requires the surprise reduction in the output gap to be $(1+\phi_s)/\phi_y$ times as large as the surprise increase in inflation, rather than only $1/\phi_y$ times as large, as under the optimal commitment assuming RE (Woodford 2003, chapter 7). Thus, as in the analysis of robustness to adaptive learning dynamics in Section 4.3.2, one again concludes that ensuring greater robustness to potential (modest) departures from fully model-consistent expectations requires one to adjust the relative weights on inflation and the output gap in a monetary policy rule in the direction of stronger relative responses to fluctuations in the rate of inflation.

6. CONCLUSION

This review illustrates only a few of the possible methods of macroeconomic analysis that depart in one way or another from the complete requirements of the RE hypothesis. Rather than presenting all the possible specifications of expectations or reviewing all the conclusions obtained using them in particular models, I have sought only to compare broad classes of approaches, which differ in the respects in which they maintain or depart from particular aspects of the knowledge assumptions maintained in the RE literature. Even this brief overview has shown that there is a considerable range of alternative approaches, leading to different conclusions about a variety of issues.

It may be asked how macroeconomic analysis can be possible with such a wide range of candidate assumptions. One answer would be that empirical studies should be undertaken to determine which of these possible specifications of subjective expectations best describe observed behavior. A few studies of that kind already exist, but the empirical literature remains at a fairly early stage. Much early work on the alternatives surveyed here has been undertaken to clarify or criticize the conceptual foundations of REE, rather than to provide a positive analysis of observed phenomena; further empirical applications are much to be desired.

Nonetheless, it is probably a mistake to suppose that empirical investigations should identify a single model of expectations that can be judged to have been historically valid and that can then be treated as the way in which expectations must be formed in the future, for purposes of counterfactual policy analyses. It is more reasonable, in my view, to search for policies that should be robust to a variety of possible specifications of expectations. Of course, it is not possible (and probably would not be desirable, even if feasible) to demand that a policy be robust to all possible views of the world; it is therefore important that macroeconomists continue to seek greater certainty about which models of the economy are more accurate. But one need not settle upon a single model specification before policy analysis is possible.

Indeed, the approaches discussed in Sections 3 and 5 above seek to define classes of reasonable specifications of expectations under a given policy regime, rather than a single correct specification, and even in the case of the econometric learning models discussed in Section 4, the identification of a best-fitting learning rule for some historical data set would better be taken as providing evidence about the types of learning rules that should be allowed for in a robustness analysis, rather than as identifying a true learning rule that can be relied upon in the future. If macroeconomic analysis is approached in this spirit, then awareness of a variety of arguably reasonable specifications should contribute to the robustness of the conclusions reached, rather than preventing any policy recommendations from being given.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

I would like to thank Klaus Adam, Ben Hebert, Mordecai Kurz, David Laibson, and Bruce Preston for helpful discussions; Savitar Sundaresan for research assistance; and the Institute for New Economic Thinking and the Taussig Visiting Professorship, Harvard University, for supporting this research.

LITERATURE CITED

Adam K, Woodford M. 2012. Robustly optimal monetary policy in a microfounded new Keynesian model. *J. Monet. Econ.* 59:468–87

Bassetto M. 2002. A game-theoretic view of the fiscal theory of the price level. *Econometrica* 70:2167–95

Benhabib J, Evans GW, Honkapohja S. 2012. *Liquidity traps and expectation dynamics: fiscal stimulus or fiscal austerity?* NBER Work. Pap. 18114

Benhabib J, Schmitt-Grohé S, Uribe M. 2001. The perils of Taylor rules. *J. Econ. Theory* 96:40–69

Bernheim D. 1984. Rationalizable strategic behavior. *Econometrica* 52:1007–28

Blanchard OJ, Kahn C. 1980. The solution of linear difference equations under rational expectations. *Econometrica* 48:1305–11

Branch WA. 2004. Restricted perceptions equilibria and learning in macroeconomics. In *Post Walrasian Macroeconomics: Beyond the Dynamic Stochastic General Equilibrium Model*, ed. D Colander, pp. 135–60. Cambridge, UK: Cambridge Univ. Press

Bray M. 1982. Learning, estimation and stability of rational expectations. *J. Econ. Theory* 26:318–39

Bullard J, Mitra K. 2002. Learning about monetary policy rules. *J. Monet. Econ.* 49:1105–29

Christiano LJ, Eichenbaum M, Evans CL. 2005. Nominal rigidities and the dynamic effects of a shock to monetary policy. *J. Polit. Econ.* 113:1–45

Clarida R, Gali J, Gertler M. 1999. The science of monetary policy. *J. Econ. Lit.* 37:1661–707

Clarida R, Gali J, Gertler M. 2000. Monetary policy rules and macroeconomic stability: evidence and some theory. *Q. J. Econ.* 115:147–80

Cover TM, Thomas JA. 2006. *Elements of Information Theory*. New York: Wiley Intersci. 2nd ed.

Deaton A. 1992. *Understanding Consumption*. New York: Oxford Univ. Press

De Grauwe P. 2010. Top-down versus bottom-up macroeconomics. *CESifo Econ. Stud.* 56:465–97

Eusepi S, Preston B. 2011. Expectations, learning and business cycle fluctuations. *Am. Econ. Rev.* 101:2844–72

Eusepi S, Preston B. 2012a. Debt, policy uncertainty, and expectations stabilization. *J. Eur. Econ. Assoc.* 10:860–86

Eusepi S, Preston B. 2012b. *Fiscal foundations of inflation: imperfect knowledge*. Unpublished manuscript, Fed. Reserve Bank New York

Evans GW, Honkapohja S. 2001. *Learning and Expectations in Macroeconomics*. Princeton, NJ: Princeton Univ. Press

Evans GW, Honkapohja S. 2003. Expectations and the stability problem for optimal monetary policies. *Rev. Econ. Stud.* 70:807–24

Evans GW, Honkapohja S. 2006. Monetary policy, expectations and commitment. *Scand. J. Econ.* 108:15–38

Evans GW, Honkapohja S. 2009. Learning and macroeconomics. *Annu. Rev. Econ.* 1:421–51

Evans GW, Honkapohja S. 2010. Expectations, deflation traps, and macroeconomic policy. In *Twenty Years of Inflation Targeting: Lessons Learned and Future Prospects*, ed. D Cobham, Ø Eitrheim, S Gerlach, JF Qvigstad, pp. 232–60. Cambridge, UK: Cambridge Univ. Press

Evans GW, Honkapohja S, Mitra K. 2012. Does Ricardian equivalence hold when expectations are not rational? *J. Money Credit Bank.* 44:1259–83

Evans GW, McGough B. 2009. *Learning to optimize*. Unpublished manuscript, Univ. Oregon, Eugene

Evans GW, Ramey G. 1992. Expectation calculation and macroeconomic dynamics. *Am. Econ. Rev.* 82:207–24

Friedman M. 1968. The role of monetary policy. *Am. Econ. Rev.* 58:1–17

Fuster A, Hebert B, Laibson D. 2011. Natural expectations, macroeconomic dynamics, and asset pricing. In *NBER Macroeconomics Annual*, ed. D Acemoglu, M Woodford, 26:1–48. Chicago: Univ. Chicago Press

Fuster A, Hebert B, Laibson D. 2012. Investment dynamics with natural expectations. *Int. J. Central Bank*. 8:243–65

Fuster A, Laibson D, Mendel B. 2010. Natural expectations and macroeconomic fluctuations. *J. Econ. Perspect*. 24:67–84

Gali J. 2008. *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*. Princeton, NJ: Princeton Univ. Press

Gaspar V, Smets F, Vestin D. 2011. Inflation expectations, adaptive learning and optimal monetary policy. In *Handbook of Monetary Economics*, Vol. 3B, ed. BM Friedman, M Woodford, pp. 1055–95. Amsterdam: Elsevier

Grandmont J-M. 1977. Temporary general equilibrium theory. *Econometrica* 45:535–72

Grandmont J-M, ed. 1988. *Temporary Equilibrium: Selected Readings*. New York: Academic

Guesnerie R. 1992. An exploration of the eductive justifications of the rational expectations hypothesis. *Am. Econ. Rev*. 82:1254–78

Guesnerie R. 2005. *Assessing Rational Expectations 2: "Eductive" Stability in Economics*. Cambridge, MA: MIT Press

Guesnerie R. 2009. Macroeconomic and monetary policies from the eductive viewpoint. In *Monetary Policy Under Uncertainty and Learning*, ed. K Schmidt-Hebbel, C Walsh, pp. 171–202. Santiago: Central Bank Chile

Hansen LP, Sargent TJ. 2008. *Robustness*. Princeton, NJ: Princeton Univ. Press

Hansen LP, Sargent TJ. 2012. *Three types of ambiguity*. Unpublished manuscript, Univ. Chicago

Harvey AC. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge, UK: Cambridge Univ. Press

Hicks J. 1939. *Value and Capital*. Oxford: Clarendon

Howitt P. 1992. Interest-rate control and nonconvergence to rational expectations. *J. Polit. Econ*. 100:776–800

Kurz M. 1994. On the structure and diversity of rational beliefs. *Econ. Theory* 4:877–900

Kurz M, ed. 1997. *Endogenous Economic Fluctuations: Studies in the Theory of Rational Belief*. Berlin: Springer

Kurz M. 2012. *A new Keynesian model with diverse beliefs*. Unpublished manuscript, Stanford Univ., Stanford, CA

Kurz M, Motolese M. 2011. Diverse beliefs and time variability of risk premia. *Econ. Theory* 47:293–335

Kwon H, Miao J. 2012. *Woodford's approach to robust policy analysis in a linear-quadratic framework*. Unpublished manuscript, Boston Univ.

Lindahl E. 1939. *Theory of Money and Capital*. London: Allen & Unwin

Lubik TA, Schorfheide F. 2004. Testing for indeterminacy: an application to U.S. monetary policy. *Am. Econ. Rev*. 94:190–217

Marcet A, Sargent TJ. 1989. Convergence of least squares learning mechanisms in self-referential linear stochastic models. *J. Econ. Theory* 48:337–68

Maskin E, Tirole J. 2001. Markov perfect equilibrium: I. Observable actions. *J. Econ. Theory* 100:191–219

McCallum BT. 1983. On nonuniqueness in linear rational expectations models: an attempt at perspective. *J. Monet. Econ*. 11:134–68

Milani F. 2005. Learning, monetary policy rules, and macroeconomic stability. *J. Econ. Dyn. Control* 32:3148–65

Milani F. 2007. Expectations, learning and macroeconomic persistence. *J. Monet. Econ*. 54:2065–82

Milani F. 2011. Expectation shocks and learning as drivers of the business cycle. *Econ. J*. 121:379–401

Muth JF. 1960. Optimal properties of exponentially weighted forecasts. *J. Am. Stat. Assoc*. 55:299–305

Orphanides A, Williams JC. 2005. Imperfect knowledge, inflation expectations and monetary policy. In *The Inflation Targeting Debate*, ed. BS Bernanke, M Woodford, pp. 201–46. Chicago: Univ. Chicago Press

Pearce D. 1984. Rationalizable strategic behavior and the problem of perfection. *Econometrica* 52:1029–50

Phelps ES. 1983. The trouble with "rational expectations" and the problem of inflation stabilization. In *Individual Forecasting and Aggregate Outcomes: "Rational Expectations" Reconsidered*, ed. R Frydman, ES Phelps, pp. 31–46. Cambridge, UK: Cambridge Univ. Press

Preston B. 2005. Learning about monetary policy rules when long-horizon expectations matter. *Int. J. Central Bank*. 1:81–126

Rondina G, Walker TB. 2012. *Information equilibria in dynamic economies with dispersed information*. Unpublished manuscript, Univ. Calif., San Diego

Sargent TJ. 1993. *Bounded Rationality in Macroeconomics*. New York: Oxford Univ. Press

Sargent TJ. 2008. Evolution and intelligent design. *Am. Econ. Rev*. 98:5–37

Slobodyan S, Wouters R. 2009. *Learning in an estimated medium-scale DSGE model*. Work. Pap. 396, CERGE-EI, Prague

Smets F, Wouters R. 2007. Shocks and frictions in US business cycles: a Bayesian DSGE approach. *Am. Econ. Rev*. 97:586–606

Svensson LEO. 1999. Inflation targeting as a monetary policy rule. *J. Monet. Econ*. 43:607–54

Taylor JB. 1993. Discretion versus policy rules in practice. *Carnegie-Rochester Conf. Ser. Public Policy* 39:195–214

Walsh CE. 2010. *Monetary Theory and Policy*. Cambridge, MA: MIT Press. 3rd ed.

Woodford M. 2001. Fiscal requirements for price stability. *J. Money Credit Bank*. 33:669–728

Woodford M. 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton, NJ: Princeton Univ. Press

Woodford M. 2010. Robustly optimal monetary policy with near-rational expectations. *Am. Econ. Rev*. 100:274–303

Annual Review of Economics

Volume 5, 2013

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Economics* articles may be found at http://econ.annualreviews.org