Are Low Interest Rates Deflationary? A Paradox of Perfect-Foresight Analysis*

Mariana García-Schmidt
Central Bank of Chile†

Michael Woodford
Columbia University‡

January 15, 2018

Abstract

We argue that an influential "neo-Fisherian" analysis of the effects of low interest rates depends on using perfect foresight equilibrium analysis under circumstances where it is not plausible for people to hold expectations of that kind. We propose an explicit cognitive process by which agents may form their expectations of future endogenous variables. Perfect foresight is justified by our analysis as a reasonable approximation in some cases, but in the case of a commitment to maintain a low nominal interest rate for a long time, our reflective equilibrium implies neither neo-Fisherian conclusions nor implausibly strong predicted effects of forward guidance.

*We would like to thank John Cochrane, Gauti Eggertsson, Jamie McAndrews, Rosemarie Nagel, Jón Steinsson, Lars Svensson, and an anonymous referee for helpful comments, and the Institute for New Economic Thinking for research support.

†Agustinas 1180, Santiago, Chile (e-mail: mcgarcia@bcentral.cl)

‡420 West 119th street, New York, NY 10027 (e-mail: michael.woodford@columbia.edu)

1 Paradoxes in Analyses of the Effects of Forward Guidance

During the global financial crisis and its aftermath, the Federal Reserve and many other central banks found their policies constrained by an effective lower bound on the level at which they could set the short-run nominal interest rates that they used as their main instruments of policy. This led to a variety of dramatic experiments with unconventional policies, intended to provide additional stimulus to aggregate demand with requiring additional reductions in short-term nominal interest rates. One such policy was "forward guidance" Woodford (2013b) — a promise to maintain unusually accommodative policy (through future use of the nominal interest-rate instrument) for a longer time than might otherwise have been expected on the basis of the bank's pre-crisis reaction function.

The effectiveness of such policies as a source of demand stimulus is a matter of debate. One skeptical argument begins by noting that many models imply that in an environment with no stochastic disturbances, a monetary policy that maintains a constant rate of inflation will also involve a constant nominal interest rate, equal to the target inflation rate plus a constant, as the steady-state real rate of interest should be independent of monetary policy. "Neo-Fisherians" propose on this ground that a central bank that wishes to bring about a given rate of inflation should commit to maintain a nominal interest rate that is higher, the higher the desired rate of inflation; but on this view, a commitment to keep nominal interest rates low should be disinflationary. Indeed, authors beginning with Bullard (2010) and Schmitt-Grohé and Uribe (2010) have proposed that a central bank faced with persistently low inflation despite a nominal interest rate at the effective lower bound should actually *raise* its interest-rate target in order to head off the possibility of a deflationary trap.

Such reasoning is quite different, of course, from that which has guided the forward guidance experiments of central banks. And theoretical analyses in the context of New Keynesian models, such as those of Eggertsson and Woodford (2003), Levin et al. (2010), and Werning (2012), find that a policy commitment that implies that the interest rate will remain at its lower bound for a few quarters longer should, if believed, result in substantially higher inflation and real activity immediately. Such analyses,

however, raise important questions.

If one considers the thought experiment of committing to keep the interest rate at its lower bound for a fixed period of time,[1] followed by immediate reversion to a policy that achieves the bank's normal inflation target, then a perfect-foresight equilibrium analysis that makes use of the equilibrium selection that is conventional in the New Keynesian literature (the forward-stable perfect-foresight equilibrium, explained below) will predict a rate of inflation and a level of real activity that are both increasing in the length of the commitment to the fixed interest rate. Moreover, the effects of the policy on these variables is not only steadily increasing: it is predicted to grow explosively, and without bound, as the horizon is lengthened. This prediction of explosive effectiveness may seem difficult to square with the modest effects of actual experiments with forward guidance — a problem that Del Negro et al. (2015) christen "the forward guidance puzzle."

And regardless of whether one regards the theoretical proposition as having been tested empirically,[2] the explosiveness result has an uncomfortable implication. If a commitment to fix the nominal interest rate for 150 years would be vastly more expansionary than a "mere" commitment to fix it for 100 years, this implies that alternative policy commitments that differ only in what is specified about policy more than a century from now should have greatly different effects; yet extreme sensitivity to changes in expectations about policy very far in the future is an unappealing feature for a model to have.

The predictions of perfect foresight analysis are even more paradoxical if one considers the more extreme (though conceptually simple) thought experiment of a *permanently* lower interest-rate peg. In this case, a standard New Keynesian model has a continuum of perfect foresight equilibria that remain forever bounded; all of them converge eventually to a steady state with a constant inflation rate lower than the central bank's inflation target, by the same number of percentage points as the nominal interest rate peg is lower than the interest rate in the steady state consistent with the

[1] This is not the policy proposed by Eggertsson and Woodford (2003); they called for a commitment to keep the interest rate at its lower bound until a price-level target path could be hit. A number of central banks have, however, implemented "date-based" policies like the experiment discussed here.

[2] Woodford (2013b) and Andrade et al. (2016) offer potential explanations for the modest effects of actual policies that propose that not everyone interpreted the central bank to have made a commitment of the kind assumed in the thought experiment discussed here.

inflation target. Thus if one supposes that one or another of these equilibria should be the outcome in the case of such a policy commitment, one would conclude (i) that the effects on both output and inflation are bounded, even though the peg lasts forever, and (ii) that at least eventually, inflation is predicted to be lower, rather than higher, because of pegging the nominal interest rate at a lower level. Both conclusions contrast sharply with what one would conclude by considering the case of a finite-duration peg, and taking the limit of the equilibrium predictions as the duration of the peg is made unboundedly long.

In fact, we believe that both the conclusion that commitment to a sufficiently long-duration period of low interest rates should be unboundedly stimulative, and the contrary conclusion that commitment to a low interest rate peg should be disinflationary, are equally unwarranted. Both conclusions result from the use of perfect foresight (or rational expectations equilibrium) analysis under circumstances in which one should not expect such an equilibrium to arise — or even for such an equilibrium to provide a reasonable approximation to the economy's actual dynamics. The concept of perfect-foresight equilibrium assumes a correspondence between what economic agents expect the economy's evolution to be and the way that it actually evolves as a result of their actions. But it is important to consider how such a correspondence can be expected to arise; we believe that it is more plausible to expect people's actual beliefs to resemble perfect-foresight beliefs under some circumstances than others. In fact, we argue that the kind of thought experiments just proposed — in which the nominal interest rate is fixed at some level for a long time, or even permanently, independently of how inflation and output may evolve — are circumstances in which a rational process of belief revision is particularly unlikely to converge quickly, or even to converge at all, to perfect foresight equilibrium beliefs.[3] This is in our view the source of the paradoxical conclusions that are obtained by assuming that the economy must follow a perfect foresight equilibrium path.

We show that the paradoxes disappear if an alternative approach is used to model the consequences of commitment to conduct monetary policy in the future according to some novel rule. Our proposal is to model the economy as being in a *temporary*

[3]The appropriateness of drawing "Neo-Fisherian" conclusions from perfect foresight analyses of the equilibria consistent with an interest-rate peg has similarly been challenged by Evans and McGough (2017), on the basis of an analysis of adaptive learning dynamics.

equilibrium with *reflective expectations* (or "reflective equilibrium" for short). By temporary equilibrium we mean, as in the approach to dynamic economic modeling pioneered by Hicks (1939), that market outcomes at any point in time result from optimizing decisions by households and firms under expectations that are specified in the model, but that need not be correct. By reflective expectations we mean expectations formed on the basis of reasoning about how the economy should evolve, under a correct understanding of the structural relations that determine market outcomes, including the central bank's commitment to follow a particular monetary policy rule in the future.

Under certain circumstances, the process of reflection that we posit, if carried far enough, will converge to a fixed point in which the temporary equilibrium outcome implied by particular expectations is exactly the path that is expected: in this limiting case, reflective equilibrium becomes a perfect foresight equilibrium.[4] When this process converges rapidly enough, it is plausible to assume that actual outcomes (resulting from some finite level of reflection) should be similar to perfect foresight equilibrium predictions. We show that this is true, in our model, if monetary policy is expected to be conducted in accordance with a Taylor rule, and the zero lower bound on interest rates is not expected ever to be a binding constraint. This result justifies the use of perfect foresight or rational expectations analysis in exercises of the kind undertaken in Woodford (2003, sec.2.4) or Galí (2015, chap. 3) using New Keynesian models similar to the one analyzed here.

But the process of reflection need not converge quickly, or even converge at all. We show that in the case of a commitment to fix the nominal interest rate for a period of time, convergence is slower than in the case of endogenous interest-rate responses of the kind called for by a Taylor rule, and is slower the longer the time for which the interest rate is expected to be fixed. In the case of a permanent interest-rate peg, the process of reflection does not converge at all, and indeed further reflection leads only to expectations still *farther* from satisfying the requirements for a perfect foresight equilibrium. Thus if the actual effect of the policy change corresponds to a reflective equilibrium with some finite level of reflection, the predictions of the perfect foresight equilibrium analysis will be less and less reliable the longer the peg is expected to last.

[4] In a stochastic generalization of the model (not taken up here), it would become a rational expectations equilibrium.

Moreover, the apparently contradictory conclusions cited above arise only from using perfect foresight analysis in cases when we should expect it to be highly inaccurate. Under the reflective equilibrium analysis with a finite level of reflection, a commitment to a lower nominal interest rate for a period of time should increase both output and inflation, but the predicted magnitude of the effect does not grow explosively as the duration of the peg is increased; it remains bounded, and is similar for all long-enough durations, including the case of a permanent peg. Neither the prediction of extreme stimulative effects, nor the prediction that a sufficiently long-lasting commitment to a low nominal interest rate should actually *reduce* inflation, is correct under this analysis.

We proceed as follows. Section 2 presents the relationships that describe a temporary equilibrium in the case of a standard New Keynesian model, under an arbitrary specification of private-sector expectations, and then defines reflective expectations. It also illustrates the application of these concepts to the simple case of a permanent commitment to a new monetary policy rule, and shows that "neo-Fisherian" conclusions are not supported. Section 3 then considers reflective equilibrium in the more complex case of commitment to a new policy for only a finite duration, when the new policy is a shift in the intercept (or implicit inflation target) of a Taylor rule, and the interest-rate lower bound never binds. Section 4 considers reflective equilibrium in the less well-behaved case of a fixed interest rate until some horizon T, and reversion to a Taylor rule thereafter, as well as the limiting case of a permanent interest-rate peg. Section 5 offers concluding reflections.

2 Reflective Equilibrium in a New Keynesian Model

In order to consider the process through which expectations are formed, and determine their degree of similarity to those that would be derived from perfect foresight, it is necessary to carefully distinguish between two aspects of the analysis of equilibrium dynamics that are typically conflated in derivations of the "equilibrium conditions" implied by New Keynesian models. These are the relations among economic variables that are implied by optimal decision making by households and firms, given their expectations about the future evolution of variables outside their control, on the one hand, and the equations that specify how expectations are formed, on the other.

Following Hicks, we refer to the first set of equations as the *temporary equilibrium* relations implied by a given model. They can be used to predict outcomes for variables such as output and inflation under any of a variety of possible assumptions about the nature of expectations.

We begin our analysis of reflective equilibrium below by specifying the temporary equilibrium relations implied by a log-linearized New Keynesian model. Stated abstractly, these are a set of equations of the form

$$x_t = \psi(\mathbf{e}_t) \tag{2.1}$$

where x_t is a finite-dimensional vector of endogenous variables determined at time t, $\mathbf{e}_t$ is an infinite-dimensional vector (an infinite sequence of vectors) specifying average expectations at time t regarding the values of the variables x_{t+j} at each of the future horizons $j = 1, 2, \ldots$, extending indefinitely into the future,[5] and $\psi(\cdot)$ is a linear operator. Temporary equilibrium relations of this form are also used in analyses of adaptive learning dynamics, such as Evans and McGough (2017). In such analyses, the model is closed by specifying a forecasting rule $\mathbf{e}_t = \phi(x_{t-1}, x_{t-2}, \ldots)$ that generates forecasts as some function of the history of past observations. The adaptive learning dynamics are then described by a dynamical system $x_t = \psi(\phi(x_{t-1}, x_{t-2}, \ldots))$.

Evans and McGough (2017) use an analysis of this kind to consider the effects of raising the level at which the short-term nominal interest rate is pegged, and show that the learning dynamics do not converge, even far in the future (assuming that the peg were to be maintained indefinitely), to the steady state with higher inflation predicted by the perfect foresight analysis. While this casts doubt on the relevance of the perfect foresight analysis (as we do), because expectation formation is based purely on past experience, their analysis addresses only the *eventual* effects of a change in policy after it has been implemented and its consequences observed for some time. It does not say anything about the effects of forward guidance; in the Evans-McGough analysis, the effects of the policy change are the same whether the new policy is announced or not.

We are concerned instead with the immediate effects of an announcement that

[5] In the model specified below, both households and firms solve infinite-horizon decision problems, and their optimal decisions depend on expectations regarding variables arbitrarily far in the future, as stressed by Preston (2005). Because the model is linearized, only the average expectations of each population of decision makers matters for aggregate outcomes.

(because the interest-rate lower bound constrains current policy) will have no consequences in the short run for observed policy. The effects of forward guidance, if any, depend on a more sophisticated approach to expectation formation, which allows announced changes in policy to be factored into the way that people think about what should happen in the future. Moreover, we wish to consider the effects of announcing a policy that may never have been tried previously, so that structural knowledge, rather than pure extrapolation from past experience, must be used to draw conclusions about the import of the announcement.

Our alternative specification of expectations is similar to the concept of "level-k reasoning" that has been argued to provide a realistic description of expectation formation in experimental games, especially when a game is played for the first time, so that expectations about other players' actions must be based on reasoning from the announced structure of the game rather than experience.[6] The "level-k" model of boundedly rational play begins with a specification of a naive approach to the game ("level-0 reasoning"), and then considers how someone should play who optimizes but assumes that the other players will play naively ("level-1 reasoning"), how someone should play who optimizes and assumes that the other players will use level-1 reasoning ("level-2 reasoning"), and so on.

In order to apply this kind of reasoning to the dynamic decision problems of our model, we need to define a mapping from a *sequence* $\mathbf{e}$ of expectations about aggregate outcomes (as a result of others' degree of reflection), extending indefinitely into the future, to a *sequence* $\mathbf{e}^*$ of outcomes, also extending indefinitely into the future, resulting from optimization given those expectations, as in the "calculation equilibrium" of Evans and Ramey (1992, 1995, 1998).[7] Below we show how the temporary equilibrium relations can be used to define such a mapping,

$$\mathbf{e}^* = \mathbf{\Psi}(\mathbf{e}). \tag{2.2}$$

Starting from any specification $\mathbf{e}(0)$ of naive expectations, one can then define "level-k

[6] See, for example, Nagel (1995), Camerer et al. (2004), Arad and Rubinstein (2012), and further discussion in García-Schmidt and Woodford (2015).

[7] See García-Schmidt and Woodford (2015) for further discussion of the relation of the Evans-Ramey "calculation equilibrium" to our own concept of reflective equilibrium.

expectations" for any finite level of reflection k by the sequence $\mathbf{e}^*(k) = \mathbf{\Psi}^k(\mathbf{e}(0))$.[8]

We do not, however, consider it most reasonable to model the result of a finite degree of reflection about the implications of structural knowledge by "level-k expectations" of this kind, for some discrete value of k. Instead, we consider a continuous process of belief revision, and define "reflective expectations" corresponding to some continuous degree of reflection n as the sequence of expectations $\mathbf{e}(n)$ given by the solution to a differential equation system

$$\dot{\mathbf{e}}(n) \;=\; \mathbf{\Psi}(\mathbf{e}(n)) \;-\; \mathbf{e}(n), \tag{2.3}$$

where the dot indicates the derivative with respect to n, and the process is integrated forward from the initial condition $\mathbf{e}(0)$ given by the naive expectations.

The solution $\mathbf{e}(n)$ to this equation corresponds to the average expectations of a population of "level-k" reasoners, with a Poisson distribution for the level of reasoning:

$$\mathbf{e}(n) \;=\; \sum_{k=0}^{\infty} e^{-n} \frac{n^k}{k!}\, \mathbf{e}^*(k).$$

Here the continuous parameter n indexes the mean level of reasoning in the population.[9] We regard this continuous specification of the consequences of a finite degree of reflection as a more realistic model of aggregate outcomes than the assumption of a population made up of entirely of decision makers of a single level of reasoning k, but who all believe that *everyone else* has a common level of reasoning $k-1$.[10]

This model of expectation formation can provide foundations for perfect foresight

[8] Following our original proposal in García-Schmidt and Woodford (2015), boundedly rational expectations have been similarly specified in New Keynesian models by Farhi and Werning (2017), Angeletos and Lian (2017), and Iovino and Sergeyev (2017).

[9] See García-Schmidt and Woodford (2015) for additional possible interpretations of the specification of reflective expectations using equation (2.3). Note that for our results below, it only matters that *average* expectations be the ones specified by $\mathbf{e}(n)$, and not that there be heterogeneity in individual beliefs of the kind described by a Poisson distribution.

[10] Note that while experimentalists have found the concept of discrete levels of reasoning useful in explaining the behavior of individual subjects in "beauty-contest" games, such studies always find that subjects exhibit several different levels of reasoning. Use of the continuous specification (2.3) also increases the range of cases in which reflective expectations converge to perfect foresight expectations as n is made large, as shown in section E of the Appendix. See García-Schmidt and Woodford (2015) and Angeletos and Lian (2017) for further discussion of the difference between discrete and continuous modeling of the degree of reflection.

equilibrium analysis, under certain circumstances. If $\mathbf{e}(n)$ converges as $n \to \infty$, then the limiting expectations $\bar{\mathbf{e}}$ must be a fixed point of the mapping $\boldsymbol{\Psi}$: $\bar{\mathbf{e}} = \boldsymbol{\Psi}(\bar{\mathbf{e}})$. This means that the sequences $\bar{\mathbf{e}}$ must represent perfect foresight equilibrium dynamics. In such a case, the predictions associated with the perfect foresight equilibrium reached in this way can be justified as the outcome of a high degree of reflection of the kind that we model.[11] We regard it as realistic to assume only a finite (and possibly rather modest) degree of reflection; but if reflective expectations converge to perfect foresight expectations sufficiently rapidly, reliance upon perfect foresight analysis might be viewed as a useful approximation. Hence we are interested, below, not only in whether reflective expectations converge, but how rapidly they converge. This turns out to depend on the kind of commitment that is made about future monetary policy.

Finally, it is important to stress that our model of reflective equilibrium represents the outcome of a process of reflection about what one should expect the future to be like, taking place at a single point in time. Even though the outcome of the analysis (the infinite-dimensional vector $\mathbf{e}(n)$) specifies *sequences* of values for the endogenous variables, extending indefinitely into the future, this should not be regarded as a prediction of our model about how the economy should evolve far into the future. Our belief-revision process starts from an initial conjecture about what "naive" expectations would be like that is made before any observation of what actually happens. Even assuming that this conjecture is initially correct at the time of the announcement of some change in policy, the expectations of naive decision makers — that we assume are based on extrapolation from past experience, rather than deduction from structural knowledge — should eventually change, after a sufficient period of observation of outcomes under the new policy. The dynamics that are projected in the reflective equilibrium calculation should not actually be realized, unless people continue to assume exactly the same "naive" expectations, despite the availability of a historical record that comes to include data from after the policy announcement. This might be true in the short run, but surely not forever.

Thus we are concerned with the *short-run* effects of a policy change that is expected to last for a significant period of time, and even a *permanent* policy change — but not

[11] Such a justification of perfect foresight or rational expectations equilibrium predictions is closely related to the "eductive justification" proposed by Guesnerie (1992,2008). The relationship of our results to Guesnerie's analysis is discussed further in García-Schmidt and Woodford (2015).

with the longer-run effects of such policy changes. Extension of the analysis to deal with that further issue would require that we add a model of adaptive learning, to model the evolution over time of "naive" expectations. Given that the details of such an extension would be relatively independent of the way that we model the revision of beliefs through the process of reflection, we do not propose one here.

2.1 The Temporary Equilibrium Relations

We make these ideas more concrete in the context of a log-linearized New Keynesian model. The model is one that has frequently been used, under the assumption of perfect foresight or rational expectations, in analyses of the potential effects of forward guidance at the interest-rate lower bound (e.g., Eggertsson and Woodford, 2003; Werning, 2012; McKay et al., 2016; Cochrane, 2017).[12] We begin by specifying the temporary equilibrium relations that map arbitrary subjective expectations into market outcomes.[13]

The economy is made up of identical, infinite-lived households, each of which seeks to maximize a discounted flow of utility from expenditure and disutility from work, subject to an intertemporal budget constraint. We assume that at any point in time t, each household formulates a spending plan for all dates $T \geq t$ so as to maximize its utility subject to its budget constraint, given subjectively expected paths for aggregate output (that determines the household's income other than from savings), inflation, and the short-term nominal interest rate. In the present exposition, we abstract from fiscal policy, by assuming that there are no government purchases, government debt, or taxes and transfers.[14]

Under a log-linear approximation, the decision rule of household i calls for real

[12] Werning (2012) and Cochrane (2017) analyze a continuous-time version of the model, but the structure of their models is otherwise the same as the model considered here.

[13] The derivation of these relations from the underlying microfoundations is explained in section B of the Appendix. The equations given here are essentially the same as those derived in Preston (2005) and used by authors such as Evans and McGough (2017) in analyses of adaptive learning. Our exposition closely follows Woodford (2013a).

[14] Woodford (2013a) shows how the temporary equilibrium framework can be extended to include fiscal variables. The resulting temporary equilibrium relations are similar, as long as households have "Ricardian expectations" regarding their future net tax liabilities.

expenditure c_t^i in the period in which the plan is made given by

$$c_t^i \;=\; (\pi^*)^{-1}(1-\beta)\hat{b}_t^i + \sum_{T=t}^{\infty} \beta^{T-t}\, \hat{E}_t^i \left\{(1-\beta)y_T \;-\; \beta\sigma\left(i_T - \pi_{T+1} - \rho_T\right)\right\}. \qquad (2.4)$$

Here b_t^i is the net real financial wealth carried into period t, y_T is aggregate output (and hence real nonfinancial income for each household) in any period $T \geq t$, i_T is the nominal interest rate between periods T and $T+1$, π_T is the aggregate rate of inflation between periods $T-1$ and T, and ρ_T is the household's rate of time preference between periods T and $T+1$[15] (allowed to be time-varying, but assumed for simplicity to be common to all households, so that we consider only aggregate shocks). The operator $\hat{E}_t^i$ indicates that the terms involving future variables are evaluated using the household's subjective expectations, which need neither be model-consistent nor common across households.

The log-linear decision rule (2.4) is obtained by log-linearizing the conditions for an optimal intertemporal plan around a stationary equilibrium in which the rate of time preference (and other exogenous disturbances) are constant over time, and the endogenous variables output, inflation, and the nominal interest rate are also constant over time; the constant inflation rate corresponds to the central bank's "normal" inflation target π^*; and the constant nominal interest rate is the one required in order for the real rate of return to coincide with the constant rate of time preference $\bar{\rho} > 0$.[16] Each of the dated variables in (2.4) and other log-linear relations below is defined as a logarithmic deviation from the variable's stationary value.[17] The parameter $0 < \beta < 1$ is the factor by which future utility flows are discounted when the rate of time preference equals $\bar{\rho}$, and $\sigma > 0$ is the household's intertemporal elasticity of substitution, evaluated at the constant expenditure plan in the stationary equilibrium. Decision rule (2.4) generalizes the "permanent-income hypothesis" formula to allow for a non-constant desired path of spending owing either to variation in the anticipated real rate of return or transitory variation in the rate of time preference.

We assume that households correctly forecast the variations in their discount rate,

[15] Section A of the Appendix lists all variables and parameters to help follow our derivations.
[16] We assume that $\pi^* > -\bar{\rho}$, so that this nominal interest rate is positive.
[17] Thus, for example, $y_t = 0$ would mean output equal to its (positive) stationary value.

so that $\hat{E}^i_t \rho_T = \rho_T$ for all $T \geq t$.[18] Collecting the expectations regarding future conditions in one term allows us to rewrite (2.4) as

$$c^i_t \;=\; (\pi^*)^{-1}(1-\beta)\hat{b}^i_t + (1-\beta)y_t \;-\; \beta\sigma i_t \;+\; \beta g_t \;+\; \beta\, \hat{E}^i_t v^i_{t+1}, \tag{2.5}$$

where

$$g_t \;\equiv\; \sigma \sum_{T=t}^{\infty} \beta^{T-t}\, \rho_T$$

measures the cumulative impact on the urgency of current expenditure of a changed path for the discount rate, and

$$v^i_t \;\equiv\; \sum_{T=t}^{\infty} \beta^{T-t}\, \hat{E}^i_t \left\{(1-\beta)y_T \;-\; \sigma(\beta i_T - \pi_T)\right\}$$

is a household-specific subjective variable.

Then defining aggregate demand y_t (which will also be aggregate output and each household's non-financial income) as the integral of expenditure c^i_t over households i, the individual decision rules (2.5) aggregate to an *aggregate demand relation*

$$y_t \;=\; g_t \;-\; \sigma i_t \;+\; e_{1t}, \tag{2.6}$$

where

$$e_{1t} \;\equiv\; \int \hat{E}^i_t v^i_{t+1}\, di$$

is a measure of average subjective expectations.

A continuum of differentiated goods are produced by Dixit-Stiglitz monopolistic competitors, who adjust their prices as in the Calvo-Yun model of staggered pricing. A fraction $1-\alpha$ of prices are reconsidered each period, where $0 < \alpha < 1$ measures the degree of price stickiness. Our version of this model differs from many textbook presentations (but follows the original presentation of Yun, 1996) in assuming that prices that are not reconsidered in any given period are automatically increased at the target rate π^*.[19] When a firm j reconsiders its price in t, it maximizes the present

[18] Expectations of future preference shocks are thus treated differently than in Woodford (2013a). The definition of the composite expectational variable, v^i_t, is correspondingly different.

[19] This allows us to assume a positive target inflation rate — important for quantitative realism —

discounted value of profits prior to the next reconsideration of its price, given its subjective expectations regarding the evolution of aggregate demand $\{y_T\}$ and of the log Dixit-Stiglitz price index $\{p_T\}$ for all $T \geq t$. A log-linear approximation to its optimal decision rule takes the form

$$p_t^{*j} \;=\; (1-\alpha\beta)\sum_{T=t}^{\infty}(\alpha\beta)^{T-t}\hat{E}_t^j[p_T + \xi y_T - \pi^*(T-t)] - (p_{t-1}+\pi^*) \qquad (2.7)$$

where p_t^{*j} is the amount by which $j's$ log price exceeds the average of the prices that are not reconsidered, $p_{t-1}+\pi^*$, and $\xi > 0$ measures the elasticity of a firm's optimal relative price with respect to aggregate demand. The operator $\hat{E}^j[\cdot]$ indicates the subjective expectations of firm j regarding future conditions.

Again, the terms on the right-hand side of (2.7) involving subjective expectations can be collected in a single term, $\alpha\beta\hat{E}_t^j p_{t+1}^{*j}$. Aggregating across the prices chosen in period t, we obtain an *aggregate supply relation*

$$\pi_t \;=\; \kappa y_t \;+\; (1-\alpha)\beta\; e_{2t} \qquad (2.8)$$

where

$$\kappa \;\equiv\; \frac{(1-\alpha)(1-\alpha\beta)\xi}{\alpha} \;>\; 0,$$

and

$$e_{2t} \;\equiv\; \int \hat{E}_t^j p_{t+1}^{*j}\, dj$$

measures average expectations of the composite variable.

We can close the system by assuming a reaction function for the central bank of the Taylor (1993) form

$$i_t = \bar{\imath}_t + \phi_\pi \pi_t + \phi_y y_t \qquad (2.9)$$

where the response coefficients satisfy $\phi_\pi, \phi_y \geq 0$. We allow for a time-varying intercept to consider the effects of announcing a transitory departure from the central bank's normal reaction function. Equations (2.6), (2.8) and (2.9) then comprise a three-equation system, that determines the temporary equilibrium values of y_t, π_t, and i_t in a given period, as functions of the exogenous disturbances $(g_t, \bar{\imath}_t)$ and subjective

while retaining a stationary equilibrium in which the prices of all goods are identical.

expectations (e_{1t}, e_{2t}).

Under our sign assumptions, these equations necessarily have a unique solution of the form (2.1) assumed above. It is useful to write this solution in the matrix form

$$x_t = Ce_t + c\omega_t, \tag{2.10}$$

where $x_t = [y_t \quad \pi_t]'$, $e_t = [e_{1t} \quad e_{2t}]'$, $\omega_t = [g_t \quad \bar{\imath}_t]'$, and the coefficient matrices C and c are defined in section B of the Appendix.

2.2 Reflective Expectations

We model a process of reflection by a decision maker who understands how the economy works — that is, who knows the temporary equilibrium relations (2.6) and (2.8) — and who also understands and believes the policy intentions of the central bank, meaning that she knows the policy rule (2.9) in all future periods. In spite of this structural knowledge, she does not know, without further reflection, what this implies about the evolution of national income, inflation, or the interest rate (unless the policy rule specifies a fixed interest rate).

Her structural knowledge can however be used to refine her expectations about the evolution of those variables. Suppose that the decision maker starts with some conjecture about the evolution of the economy summarized by $\{e_t\}$ for each of the dates $t \geq 0$. If she assumed that others were sophisticated enough to have exactly these expectations (on average), she can then ask: What path for the economy should *she* expect, given her structural knowledge, given others' average expectations?

To answer this, we need to compute the "correct" value for the subjective expectations, $\{e^*_{1t}, e^*_{2t}\}$ given the values of the endogenous variables that are calculated using the initial expectations $\{e_{1t}, e_{2t}\}$. From the definitions of the expectational variables, e_{it}, for $i = 1, 2$, their correct values are

$$e^*_{it} = (1 - \delta_i) \sum_{j=0}^{\infty} \delta_i^j \bar{E}_t a_{i,t+j+1}, \tag{2.11}$$

where the discount factors are given by

$$\delta_1 = \beta, \qquad \delta_2 = \alpha\beta$$

so that $0 < \delta_i < 1$ for both variables,

$$\begin{aligned} a_{1t} &\equiv y_t - \frac{\sigma}{1-\beta}(\beta i_t - \pi_t), \\ a_{2t} &\equiv \frac{1}{1-\alpha\beta}\pi_t + \xi y_t, \end{aligned}$$

and $\bar{E}_t[\cdot]$ indicates the average of the population's forecasts at date t.[20]

Using (2.9) and (2.10) to substitute for i_t, π_t, and y_t in the above equations, this solution can be written in the form

$$a_t = M\, e_t + m\, \omega_t, \tag{2.12}$$

where a_t is the vector (a_{1t}, a_{2t}), and the matrices of coefficients are given in section B of the Appendix. Under any conjecture $\{e_t\}$, the temporary equilibrium relations imply unique paths for the variables $\{a_t\}$, given by (2.12). From these, the decision maker can infer implied paths $\{e_t^*\}$ for all $t \geq 0$, using equations (2.11).

If we let $\mathbf{e}^*$ and $\mathbf{e}$ denote the infinite-dimensional vectors each containing the entire sequence for $t \geq 0$ of the respective expectational variables, then we have shown how to compute all of the elements of $\mathbf{e}^*$ given a specification of the elements of $\mathbf{e}$; this defines the mapping $\mathbf{\Psi}$ introduced in (2.2).[21] We then propose that the conjectured beliefs should be adjusted in the direction of the discrepancy between the model prediction given the conjectured beliefs and the conjectured beliefs themselves, as specified in (2.3). Writing this out more explicitly, we consider a process of belief revision described by a differential equation for each date $t \geq 0$,

$$\dot{e}_t(n) = e_t^*(n) - e_t(n). \tag{2.13}$$

[20] While we still allow for the possibility of heterogeneous forecasts, from here on we simplify notation by assuming that the distribution of forecasts across households is the same as across firms.

[21] The operator Ψ also depends on the sequences of perturbations $\{\omega_t\}$, omitted to simplify notation. We apply this operator to different conjectured beliefs $\{e_t\}$, holding fixed the fundamentals.

Here the continuous variable $n \geq 0$ indexes how far the process of reflection has been carried forward, $e_t(n)$ is the conjecture of average beliefs in t at stage n, $e_t^*(n)$ is the correct forecast in period t defined by (2.11) if average expectations are given by $e_t(n)$, and $\dot{e}_t(n)$ is the derivative of $e_t(n)$ with respect to n.

We suppose that the process of reflection starts from some initial "naive" conjecture about average expectations $e_t(0)$, and that (2.13) are then integrated forward. This initial conjecture might be based on the forecasts that *would* have been correct, but for the occurrence of the unusual shock and/or change in policy that caused the process of reflection about what to expect in light of the new circumstances. The process of belief revision might be integrated forward to an arbitrary extent, but like Evans and Ramey (1992, 1995, 1998), we suppose that it would typically be terminated at some finite stage n, even if $\{e_t^*(n)\}$ still differs from $\{e_t(n)\}$.

The sequence of outcomes for $t \geq 0$ implied by the temporary equilibrium relations when average subjective expectations are given by $e_t(n)$ constitutes a *reflective equilibrium of degree n*. This will depend, of course, both on the initial expectations $e(0)$ from which the process of reflection is assumed to start, and on the stage n at which the process of reflection is assumed to terminate. Nonetheless, if the dynamics (2.13) converge globally (or at least for a large enough set of possible initial conditions) to a particular perfect foresight equilibrium, and furthermore converge rapidly enough, then a reasonably specific prediction will be possible under fairly robust assumptions. This is the case in which it would be a good approximation to use that perfect foresight equilibrium as a prediction for what should happen under the policy commitment in question. We show below that this is true when policy conforms to a Taylor rule.

But even when reflective expectations do not converge rapidly and our analysis provides only qualitative predictions, these may be of interest, since in some cases *all* of the possible outcomes are quite different from any of the perfect foresight paths. This is what we find in the case of a commitment to a fixed interest rate for a long period.

2.3 A Simple Illustration

Here we illustrate our concept of reflective equilibrium, and how it may or may not converge to a perfect foresight equilibrium in a simple special case. We consider a

stationary environment, in which $g_t = 0$ for all t, and the monetary policy reaction function (2.9) is also fixed from $t = 0$ onward ($\bar{\imath}_t = \bar{\imath}$ for all t). We further restrict attention, for this subsection only, to conjectures about others' expectations in which expectations are the same for all future periods. Thus we suppose that the average expectations at time t of the values of the variables $(\pi_{t+j}, y_{t+j}, i_{t+j})$ are given by constants (π^e, y^e, i^e) for all $j \geq 1$. In this case the expectational terms e_{it} in the temporary equilibrium relations are also the same for all t, and the temporary equilibrium relations, (2.6) and (2.8), imply that π_t, y_t, i_t will have constant values (π, y, i), given by the solutions to

$$y \;=\; -\sigma i \,+\, e_1 \tag{2.14}$$

$$\pi \;=\; \kappa y \,+\, (1-\alpha)\beta e_2 \tag{2.15}$$

with

$$e_1 \;=\; y^e \,-\, \frac{\sigma}{1-\beta}(\beta i^e - \pi^e) \qquad e_2 \;=\; \frac{\pi^e}{1-\alpha\beta} + \xi y^e$$

claculated using their definitions. The monetary policy reaction function is assumed to be:

$$i \;=\; \bar{\imath} + \phi\pi, \tag{2.16}$$

where the intercept $\bar{\imath}$ determines the implicit inflation target. We can eliminate y from (2.14)–(2.15) to yield another static relationship between π and i,

$$\pi \;=\; -\kappa\sigma i \,+\, \eta, \tag{2.17}$$

where

$$\eta \;\equiv\; \kappa e_1 + (1-\alpha)\beta e_2 = \eta_y y^e \,+\, \eta_\pi \pi^e \,+\, \eta_i i^e$$

is a composite of average expectations, with weights

$$\eta_y \;=\; \frac{\kappa}{1-\alpha\beta} \;>\; 0, \qquad \eta_\pi \;=\; \frac{\kappa\sigma}{1-\beta} + \frac{(1-\alpha)\beta}{1-\alpha\beta} \;>\; 0, \qquad \eta_i \;=\; -\frac{\beta}{1-\beta}\kappa\sigma \;<\; 0.$$

The temporary equilibrium values (π, i) are then jointly determined by the equation system (2.16)–(2.17), for given expectations η. These two equations are graphed by the lines MP (black dashed) and TE (yellow) respectively in the left panel of Figure 1; point E represents the temporary equilibrium (π, i).

A stationary temporary equilibrium of this kind will be a stationary *perfect foresight* equilibrium if expectations are correct: that is, if in addition $y^e = y, \pi^e = \pi$, and $i^e = i$. Substitution of these assumptions into (2.14) and replacing e_1, yields the *Fisher equation*

$$i \;=\; \pi \tag{2.18}$$

as another stationary relationship between i and π. A stationary perfect foresight equilibrium is then a pair of constant values (π, i) satisfying (2.16) and (2.18); this corresponds to the intersection between the lines MP and FE (dash-dotted blue) in the left panel of Figure 1. Such a situation is *also* a stationary temporary equilibrium, in which expectations η are such as to make the MP and TE curves intersect at a point on the Fisher equation locus FE, as is the case of point E in the figure.

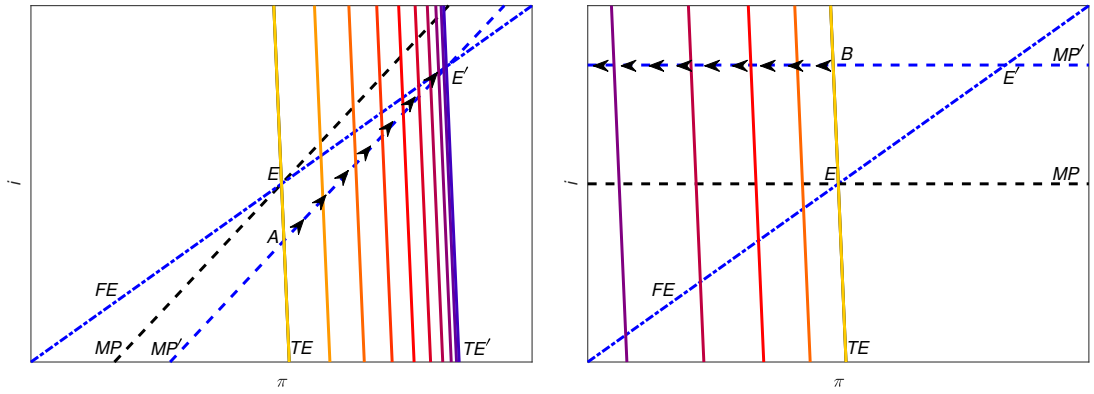
Consider a monetary policy (2.16) with $\phi > 1$, in accordance with the advice of Taylor (1993), and suppose that the policy rule has remained stable for a long enough time for people's expectations to have come to coincide with what actually occurs, so that the economy begins in a stationary perfect foresight equilibrium of the kind shown by point E in the first panel of Figure 1. (In the figure, line MP is steeper than FE because $\phi > 1$.) But consider now what should happen if the central bank announces a permanent shift to a new monetary policy, corresponding to a lower intercept $\bar{\imath}$, but the same value of ϕ. This amounts to an increase in the implicit inflation target, and shifts the MP locus down and to the right, to a new line such as MP'. Under a perfect foresight analysis, the new stationary equilibrium should be given by point E', the intersections between lines MP' and FE. This involves new stationary values for π and i that are higher by exactly the same number of percentage points.

Our reflective equilibrium analysis explains how such a new situation can be reached, rather than simply assuming that expectations must again be correct under the new policy. Suppose that naive agents (level-zero reflection) continue to expect the same paths for output, inflation and interest rates as under the previous policy rule. Thus for $n = 0$, the temporary equilibrium locus continues to be given by the line TE, despite the announcement of a new monetary policy, and the reflective equilibrium of degree $n = 0$ will be given by point A, where MP' intersects TE.

The adjustment of expectations under the process of reflection is given by

$$\dot{e} \;=\; e^* - e \;=\; [M - I](e - \bar{e}), \tag{2.19}$$

Figure 1: Temporary equilibrium determination with a stationary monetary policy.



Notes: Left panel shows the effects of a shift in the intercept of an "active" Taylor rule; right panel the effects of a shift in the level of an interest-rate peg. See text for explanation.

where the vector e indicates the time-invariant values for (e_{1t}, e_{2t}); e^* is the vector of correct values for these two expectational variables, at any given degree of reflection n; $\bar{e}$ is the value of the vector e at the stationary equilibrium represented by point E'; the dot indicates the derivative with respect to increases in n; and M is the matrix introduced in (2.12). Pre-multiplying this equation by the vector $(\kappa \ (1-\alpha)\beta)$, we obtain

$$\dot{\eta} \ = \ \eta^* - \eta \ = \ \eta_y(y - y^e) \ + \ \eta_\pi(\pi - \pi^e) \ + \ \eta_i(i - i^e), \tag{2.20}$$

where η^* is the value of η that would correspond to correct expectations in this temporary equilibrium resulting from average expectations η. This indicates the direction of shift of the expectational term in (2.17).

In the temporary equilibrium represented by point A, inflation and output are higher, and the interest rate is lower, than at point E. Thus when $n = 0$, actual outcomes are at point A while expectations are consistent with point E; hence $y > y^e, \pi > \pi^e$, and $i < i^e$. It follows from (2.20) that $\dot{\eta} > 0$, meaning that the process of reflection will shift the TE curve up and to the right in the figure. This adjustment of the temporary equilibrium locus is shown by the family of progressively darker-colored lines parallel to TE in the figure.

The adjustment dynamics are subsequently determined by (2.19).[22] We show in

[22]The simpler version (2.20) suffices to determine the sign of $\dot{\eta}$ when $n = 0$, but not once expecta-

section C of the Appendix that if $\phi > 1$, both eigenvalues of $M - I$ have negative real part, so that the dynamics implied by (2.19) converge globally to the fixed point $\bar{e}$. Thus as n increases, the temporary equilibrium locus continues to shift, converging eventually to the line labeled TE' as $n \to \infty$.[23]

The reflective equilibrium for any value of n is given by the intersection between the shifted TE locus and the line MP'; as n increases, this point moves up the MP' line from point A, as indicated by the arrows. In the limit as $n \to \infty$, the reflective equilibrium approaches point E', the stationary perfect foresight equilibrium consistent with the new monetary policy commitment. Thus our reflective equilibrium analysis explains how a commitment to a new monetary policy results in an immediate jump to a new stationary equilibrium, if the policy is believed and understood, and people are capable of a sufficient degree of reflection about its implications.[24]

The "neo-Fisherian" thesis observes that in such an equilibrium, the nominal interest rate should immediately jump to a permanently higher level, associated with a permanent increase in inflation, and argues that it should be possible to reach this new stationary equilibrium by simply announcing that the central bank will immediately *raise the nominal interest rate* to that level. Under a perfect foresight analysis, the new stationary equilibrium associated with a commitment to peg the nominal interest rate at a higher level, regardless of economic conditions (a reaction function indicated by the horizontal line MP' in the right panel of Figure 1), is the same as the one that results from a commitment to a Taylor rule of the kind indicated by the line MP' in the left panel of the figure: in each case, the stationary equilibrium is point E', the point at which the monetary policy reaction function intersects line FE.

Instead, a commitment to peg the nominal interest rate at a higher level has very different implications under the reflective equilibrium analysis. If one starts from naive expectations consistent with a previous stationary equilibrium E, then the temporary equilibrium locus corresponding to level-zero reflection is again the line TE, and under the new monetary policy MP', the reflective equilibrium of degree $n = 0$ will be given by point B in the right panel of the figure: increasing the nominal interest rate lowers

tions are no longer consistent with the former steady state E.

[23] Because the dynamics of η represent a projection onto a line of the dynamics (2.19) in the plane, the convergence need not be monotonic for all parameter values; however, convergence to the expectations represented by TE' is guaranteed.

[24] This is a special case of the more general result stated as Proposition 1 below.

both inflation and output relative to their previous values at point E. Thus we have $y < y^e, \pi < \pi^e$, and $i > i^e$, and (2.20) implies that $\dot{\eta} < 0$. So, the process of reflection will reduce η, shifting the temporary equilibrium locus down and to the left.

Moreover, in this case (or any case in which $\phi < 1$), we show in section C of the Appendix that M has a positive real eigenvalue. The belief-revision dynamics (2.19) therefore diverge from any neighborhood of the fixed point $\bar{e}$, for almost all initial conditions, including those corresponding to point B. The belief revision shifts the temporary equilibrium locus farther away from passing through the point E' (corresponding to beliefs $\bar{e}$); along the MP' line, as shown by the arrows. Reflective expectations do not converge to perfect foresight beliefs. Further reflection only makes the policy more deflationary and more contractionary. While it is true that the perfect foresight outcome E' would be a fixed point under the belief revision dynamics that we propose, it is an *unstable* fixed point — starting from any other initial conjecture leads to expectations progressively farther from it — and hence one should not expect that outcome, or one like it, to occur.

Our demonstration above of convergence to perfect foresight under a Taylor rule relies on assuming expectations of an especially simple (time-invariant) sort. Showing that the conclusion is robust requires us to consider belief revision when expectations are represented by infinite sequences. This will also allow us to consider policy commitments that apply only for particular lengths of time, as in "date-based" forward guidance. We again begin by considering policy commitments that shift the intercept of the monetary policy reaction function, but allow the shift to be time-dependent.

3 Reflective Equilibrium with a Commitment to Follow a Taylor Rule

We now consider the case in which monetary policy follows a Taylor-type rule of the form (2.9), with constant response coefficients ϕ_π, ϕ_y, but allowing a time-varying path for the intercept $\{\bar{\imath}_t\}$, here assumed to be specified by an advance commitment. The varying intercept allows us to analyze a commitment to temporarily "looser" policy, before returning to the central bank's normal reaction function. We further assume

in this section that the response coefficients satisfy

$$\phi_\pi + \frac{1-\beta}{\kappa}\phi_y \;>\; 1 \tag{3.1}$$

in conformity with the "Taylor Principle" (Woodford, 2003, chap. 4).[25]

We also assume in this section that the interest-rate lower bound never binds, so that the central bank's interest-rate target satisfies (2.9) at all times. This is an assumption that both the disturbances to fundamentals $\{\omega_t\}$ and subjective beliefs $\{e_t(n)\}$ involve small enough departures from the long-run steady state. We defer until the next section consideration of the case in which the lower bound may constrain policy for some period of time, owing to a larger shock.

We begin by recalling the perfect-foresight analysis of this case. As shown in section B of the Appendix, under the assumption of perfect foresight, (2.6) and (2.8) imply that the paths of output, inflation and the interest rate must satisfy

$$y_t = y_{t+1} - \sigma(i_t - \pi_{t+1} - \rho_t) \tag{3.2}$$

$$\pi_t = \kappa y_t + \beta \pi_{t+1} \tag{3.3}$$

which are simply perfect-foresight versions of the usual "New Keynesian IS curve" and "New Keynesian Phillips curve" respectively. Thus a *perfect foresight equilibrium* is a set of sequences $\{y_t, \pi_t, i_t\}$ that satisfy (2.9), (3.2) and (3.3) for all t, given the paths of the exogenous variables $\{\rho_t, \bar{\imath}_t\}$.

We show in section C of the Appendix that when (3.1) is satisfied, this system of equations has a unique bounded solution in the case of any bounded paths for the exogenous variables, of the form

$$x_t \;=\; \sum_{j=0}^{\infty} \zeta_j \left(\rho_{t+j} - \bar{\imath}_{t+j}\right), \tag{3.4}$$

where $\{\zeta_j\}$ converge at an exponential rate to zero for large j. We shall call this solution the "forward-stable perfect foresight equilibrium" (FS-PFE).[26] Here we show

[25] This generalizes the $\phi > 1$ case of section 2.3 to allow cases in which $\phi_y > 0$.

[26] The qualification is intended to distinguish this solution from other, explosive sequences that also satisfy equations (2.9), (3.2) and (3.3) for all t. The emphasis of the New Keynesian literature

that under certain conditions, reflective equilibrium converges to this perfect foresight equilibrium when the degree of reflection is high enough.

3.1 Exponentially Convergent Belief Sequences

Our results on the convergence of reflective equilibrium as the degree of reflection increases depend on starting from an initial ("naive") conjecture that is sufficiently well-behaved as forecasts far into the future are considered. We shall say that a sequence $\{z_t\}$ defined for all $t \geq 0$ "converges exponentially" if there exists a finite date $\bar{T}$ (possibly far in the future) such that for all $t \geq \bar{T}$, the sequence is of the form

$$z_t \;=\; z_\infty + \sum_{k=1}^{K} u_k \lambda_k^{t-\bar{T}}, \tag{3.5}$$

where z_∞ and the $\{u_k\}$ are a finite collection of real coefficients, and the $\{\lambda_k\}$ are real numbers satisfying $|\lambda_k| < 1$. This places no restrictions on the sequence over any finite time horizon, only that it converges to its long-run value in a sufficiently regular way. We shall similarly say that a vector sequence, such as $\{e_t\}$, converges exponentially if this is true of each of the individual sequences.

We shall consider only the case in which the initial belief sequence $\{e_t(0)\}$ converges exponentially. This is not motivated by any assumption that people should believe on theoretical grounds that the economy's dynamics ought to be convergent; the initial conjecture is refined (through the belief-revision process) using structural knowledge about inflation and output determination, but is not itself already based on such knowledge. Instead, the initial conjecture is intended to represent expectations that people would reasonably hold on the basis of a purely atheoretical extrapolation of past experience. The assumption of exponential convergence reflects the idea that people forming atheoretical forecasts of this kind have little reason to make different forecasts for different dates far in the future; hence as t becomes large, their forecasts converge to constant "long-run" forecasts of output and inflation (though these need not be correct). Forecasts generated by stationary ARMA models estimated using past data would be an example of an initial conjecture of this kind.[27]

on the FS-PFE has been criticized by authors such as Cochrane (2011).

[27] Of course, a different kind of initial conjecture could make sense if the situation in which the policy

Note that we do not assume that the initial "naive" conjecture necessarily converges to the long-run equilibrium values consistent with the announced policy rule. Indeed, in the case of a permanent policy change, it is most natural to assume that it does not. The initial conjecture should more likely be that what people expect in the long run is based on some average of values observed prior to the policy change, which are unrelated to what will happen under the new policy.

Note also that the temporary equilibrium relations (2.11)–(2.12) imply that if the sequence of fundamentals $\{\omega_t\}$ and a conjecture $\{e_t\}$ regarding average subjective expectations converge exponentially, then the correct expectations $\{e_t^*\}$ also converge exponentially. Thus the operator Ψ maps exponentially convergent belief sequences into exponentially convergent belief sequences. An initial conjecture that is exponentially convergent will then result in reflective expectations that are also exponentially convergent, for any order n of reflection.

3.2 Reflection Dynamics

We now consider the adjustment of the sequence $\{e_t(n)\}$ describing subjective beliefs as the process of reflection specified by (2.13) proceeds (as n increases), assuming that fundamentals $\{\omega_t\}$ and the initial conjecture $\{e_t(0)\}$ converge exponentially.[28] This allows, but does not require, the "naive" hypothesis that people's expectations are unaffected by either the shock or the change in policy that has occurred. In this case we have our first proposition:

Proposition 1 *Consider the case of a shock sequence $\{g_t\}$ that converges exponentially, and let the forward path of policy be specified by a sequence of reaction functions (2.9), where the coefficients (ϕ_π, ϕ_x) are constant over time and satisfy (3.1), and the sequence of perturbations $\{\bar{\imath}_t\}$ converges exponentially. Then in the case of any initial conjecture $\{e_t(0)\}$ regarding average expectations that converges exponentially, the belief revision dynamics (2.13) converge as n grows without bound to the belief sequence $\{e_t^{PF}\}$ associated with the FS-PFE.*

change occurs were one in which inflation and the output gap have not been relatively stationary over the recent past. The analysis in this paper applies to forward guidance experiments in situations where these variables have been relatively stationary prior to some disturbance that motivates the policy change, as during the "Great Moderation" period preceding the financial crisis of 2007-08.

[28] For a more detailed explanation refer to García-Schmidt and Woodford (2015, sec. 3.2).

The implied reflective equilibrium paths for output, inflation and the nominal interest rate similarly converge to the FS-PFE paths for these variables. This means that for any $\epsilon > 0$, there exists a finite $n(\epsilon)$ such that for any degree of reflection $n > n(\epsilon)$, the reflective equilibrium value will be within a distance ϵ of the FS-PFE prediction for each of the three variables and at all horizons $t \geq 0$.

Further details of the proof are given in section D of the Appendix.

This result has several implications. First, it shows how a perfect foresight equilibrium can arise through a process of reflection of the kind proposed in section 2. Thus it provides an explanation for treating the FS-PFE paths as the model's prediction for the effects of such a policy commitment.[29] Proposition 1 also shows that one can sometimes obtain quite precise predictions from the hypothesis of reflective equilibrium. In the case considered here, the reflective equilibrium predictions are quite similar, for all sufficiently large values of n regardless of the initial conjecture that is assumed, as long as the initial conjecture is sufficiently well-behaved. Finally, we see that the FS-PFE provides a useful approximation of the reflective equilibrium with a greater accuracy the greater the degree of reflection.

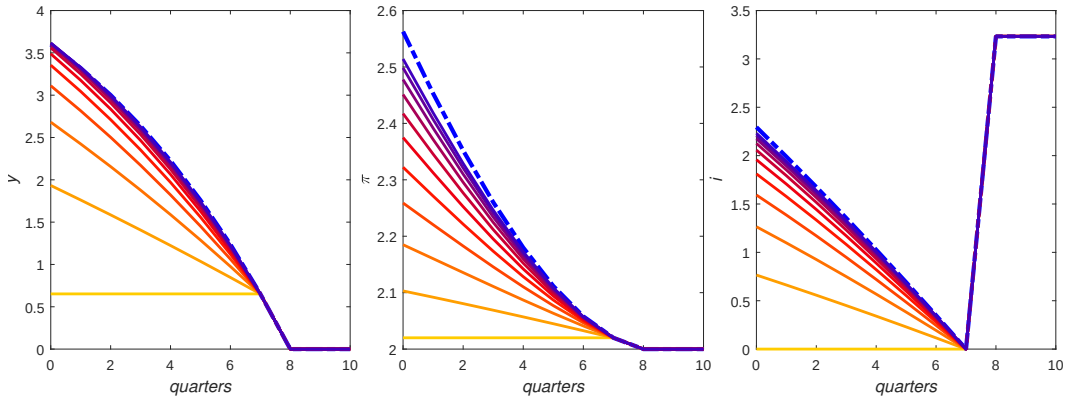
How large n must be for reflective equilibrium to resemble the FS-PFE will depend on parameter values. At least in some cases, the required n may not be implausibly large. We illustrate this with a numerical example.

Figure 2 considers an experiment in which the intercept $\bar{\imath}_t$ is lowered for 8 quarters (periods $t = 0$ through 7), but is expected to return to its normal level afterward. The policy to which the central bank returns in the long run is specified in accordance with Taylor (1993): the implicit inflation target π^* is 2 percent per annum, and the reaction coefficients are $\phi_\pi = 1.5, \phi_y = 0.5/4$.[30] The model's other structural parameters are those used by Denes et al. (2013), to show that the ZLB can produce a contraction similar in magnitude to the U.S. "Great Recession," in the case of a shock to the path of $\{g_t\}$ of suitable magnitude and persistence: $\alpha = 0.784$, $\beta = 0.997$, $\sigma^{-1} = 1.22$, and

[29] Other perfect foresight solutions can also be obtained if we assume an initial conjecture of a sufficiently different type. In particular, any perfect foresight equilibrium can be obtained if one starts from an initial conjecture given by exactly that path. However, the FS-PFE is not just a possible limit point of such a process, but a *stable* limit point, in the sense that reflective equilibrium converges to it from any of a broad range of possible initial conjectures.

[30] The division of ϕ_y by 4, relative to the value quoted by Taylor (1993), reflects the fact that periods in our model are quarters, so that i_t and π_t in (2.9) are quarterly rates.

Figure 2: Reflective equilibrium for short term change in Taylor rule intercept



Notes: The graph shows the outcomes for $n = 0$ through 4 (progressively darker lines) compared with the FS-PFE solution (dash-dotted line), when the Taylor-rule intercept is reduced for 8 quarters. See section F of the Appendix for details.

$\xi = 0.125$.[31] These imply a long-run steady-state value for the nominal interest rate of 3.23 percent per annum.[32]

We assume that $\bar{\imath}_t$ is reduced by 0.008 in quarterly units for the first 8 quarters; this is the maximum size of policy shift for which the ZLB does not bind in reflective equilibria associated with any $n \geq 0$.[33] In computing the reflective equilibria shown in Figure 2, we assume as initial "naive" conjecture the one that is correct in the steady state with 2 percent inflation. Finally, for simplicity we consider a pure temporary loosening of monetary policy, not motivated by any real disturbance (so that $g_t = 0$ for all t).[34]

[31]The parameters in Denes et al. (2013) are of interest as a case in which an expectation of remaining at the ZLB for several quarters has very substantial effects under rational-expectations.

[32]The intercept of the central-bank reaction function assumed in the long run is smaller than in Taylor (1993) and is consistent with the 2% inflation target in the steady state.

[33]As shown in Figure 2, the shock results in a zero nominal interest rate in each of the first 8 quarters, when $n = 0$. In quarter 7, the nominal interest rate is also zero for all $n \geq 0$ (and also in the FS-PFE), since expectations do not change from $t = 8$ onward.

[34]Because our model is linear, we can separately compute the perturbations implied by a pure monetary policy shift, by a real disturbance, and by a change in the initial conjecture, and sum these to obtain the predicted effects of a scenario under which a real disturbance provokes both a change in monetary policy and in the initial conjecture.

The three panels of the Figure show the temporary equilibrium paths of output, inflation and the nominal interest rate,[35] in reflective equilibria corresponding to successively higher degrees of reflection. The lightest of the solid lines (most yellow, if viewed in color) corresponds to $n = 0$; these are the outcomes that are expected to occur under the "naive" conjecture about average expectations, but taking account of the announced change in the central bank's behavior. Thus the $n = 0$ lines represent the paths that it would be correct to expect, if people all hold the initial "naive" beliefs. The reduction in the interest rate has some stimulative effect on output even in the absence of any change in expectations, but this effect is the same in each of the first 8 quarters.

As n increases, the effects on output and inflation become greater in quarters zero through 6; and the extent to which this is so is greater, the larger the number of quarters for which the looser policy is expected to continue. There are no changes in the expected paths from quarter 8 onward, since we have assumed reversion to the long-run steady-state policy in quarter 8, and the initial conjecture already corresponds to a perfect foresight equilibrium. There are similarly no changes in the expected outcomes in quarter 7, because quarter 7 *expectations* about later quarters do not change. However, the fact that outcomes are different in quarter 7 and earlier than those anticipated under the "naive" expectations causes beliefs to be revised in quarters 6 and earlier. As expectations shift toward expecting higher output and inflation, the temporary equilibrium levels of output and inflation in the earlier quarters increase (and the nominal interest rate increases as well, through an endogenous policy reaction).

The progressively darker solid lines in the figure plot the reflective equilibrium outcomes for degrees of reflection $n = 0, 0.4, 0.8$, and so on up to $n = 4.0$. The FS-PFE paths are also shown by dark dash-dotted lines. One sees that the reflective equilibrium paths converge to the FS-PFE solution as n increases, in accordance with Proposition 1. Moreover, the convergence is relatively fast. Already when $n = 2$, the predicted reflective equilibrium responses for both output and inflation differ from the perfect foresight responses by less than 10 percent in any quarter. This means that

[35] Here y_t is measured in percentage points of deviation from the steady-state level of output: for example, "2" means 2 percent higher than the steady-state level. The variables π_t and i_t are reported as annualized rates, in percentage points.

if the average member of the population is expected to be capable of iterating the Ψ mapping at least twice,[36] one should predict outcomes approximately the size of the perfect foresight equilibrium outcomes. When $n = 4$, the reflective equilibrium output responses differ from the perfect foresight outcomes by only 1 percent or less, and except in quarter zero (when the discrepancy is closer to 2 percent), the same is true of the inflation responses.

3.3 Effects of Announcing a Long-Lasting Policy Change

The paradoxes discussed in section 1 concern the predicted responses, under perfect foresight analysis, to policy changes that are expected to last for a long period of time, or even forever. Here we show that if the policy change is simply a shift in the Taylor-rule intercept, no paradoxes arise, regardless of the duration of the commitment to a different intercept (which may be understood as a temporarily different inflation target).

For the sake of specificity, we consider the following special class of policy experiments. Suppose that $\bar{\imath}_t$ is expected to take one value ($\bar{\imath}_{SR}$) for all $t < T$, and another value ($\bar{\imath}_{LR}$) for all $t \geq T$, while $g_t = 0$ for all t. (The policy experiment considered in Figure 2 was one of this kind, with $T = 8$.) We again assume that both $\bar{\imath}_{SR}$ and $\bar{\imath}_{LR}$ are high enough that the interest-rate lower bound never binds. We wish to consider how the effect of such a policy commitment depends on the horizon T. In particular, we are interested in whether the effect is similar for all large enough values of T, in order to avoid the paradoxical conclusion that alternative commitments that differ only in what the central bank promises to do very far in the future can have significantly different effects now.

Again we consider the effects of a pure policy change and an initial "naive" conjecture in which average expectations are consistent with the steady state in which the inflation target π^* is achieved at all times. Since our model is purely forward-looking, and $\bar{\imath}_t$ and $e_t(0)$ are each the same for all $t \geq T$, the belief-revision dynamics (2.13) result in $e_t(n)$ having the same value for all $t \geq T$. Let this value be denoted $e_{LR}(n)$.

[36] As discussed in García-Schmidt and Woodford (2015), the value of n in our model can be interpreted as the mean number of times that different members of the population iterate the Ψ mapping. Observed play in experimental games often suggests that a value around 2 is realistic. See, e.g., Camerer et al. (2004) and Arad and Rubinstein (2012).

It evolves according to the differential equation (2.19) introduced in our discussion of a stationary policy, starting from the initial condition $e_{LR}(0) = 0$. In this equation, the expectations in the perfect foresight steady state consistent with the LR policy rule are given by

$$\bar{e}_{LR} \equiv [I - M]^{-1} m_2 \bar{\imath}_{LR}, \tag{3.6}$$

where m_2 is the second column of the matrix m in (2.12). Here we assume that the quantity on the left-hand side of (3.1) is not exactly equal to 1,[37] in which case we can show (see the section C of the Appendix) that $M - I$ is non-singular, and (3.6) is well-defined.

The differential equation (2.19) can be solved in closed form (as discussed in section B of the Appendix) for $e_{LR}(n)$. If the reaction coefficients (ϕ_π, ϕ_y) satisfy the Taylor Principle (3.1), then as noted in section 2.3, we can show that

$$\lim_{n \to \infty} e_{LR}(n) = \bar{e}_{LR}. \tag{3.7}$$

Thus the reflective equilibrium in any period $t \geq T$ converges to the perfect foresight steady state associated with the long-run policy (which is also the FS-PFE solution for this policy). This is as we should expect from Proposition 1.

We turn now to the characterization of reflective equilibrium in periods $t < T$. The forward-looking structure of the model similarly implies that the solution for $e_t(n)$ depends only on how many periods prior to period T the period t is, and not on either t or T. If we adopt the alternative numbering scheme $\tau \equiv T - t$, then the solution for $e_\tau(n)$ for any $\tau \geq 1$ will be independent of T. We show in section B of the Appendix that we can solve the differential equation implied by the belief-revision dynamics for $e_\tau(n)$ when $\tau = 1$, using the already computed solution for $e_{LR}(n)$; then use the solution for the case $\tau = 1$ to solve for $e_\tau(n)$ when $\tau = 2$; and so on, recursively, for progressively higher values of τ.

Considering how $e_t(n)$ changes (for any fixed t) as T is increased is equivalent to considering how the solution $e_\tau(n)$ changes for larger values of τ. In particular, the behavior of $e_t(n)$ as T is made unboundedly large can be determined by calculating

[37] This condition is satisfied by generic reaction functions of the form (2.9) whether the Taylor Principle is satisfied or not. Hence we do not discuss the knife-edge case in which $M - I$ is singular, though our methods can easily be applied to that case as well.

the behavior of $e_\tau(n)$ as $\tau \to \infty$. This yields the following simple result.

Proposition 2 *Consider the case in which $g_t = 0$ for all t, and let the forward path of policy be specified by a sequence of reaction functions (2.9), where the coefficients (ϕ_π, ϕ_x) are constant over time and such that the left-hand side of (3.1) is not equal to one, and suppose that $\bar{\imath}_t = \bar{\imath}_{SR}$ for all $t < T$ while $\bar{\imath}_t = \bar{\imath}_{LR}$ for all $t \geq T$. Then if the initial conjecture is given by $e_t(0) = 0$ for all t, the reflective equilibrium beliefs $\{e_t(n)\}$ for any degree of reflection n converge to a well-defined limiting value*

$$e_{SR}(n) \equiv \lim_{T \to \infty} e_t(n)$$

that is independent of t, and this limit is given by

$$e_{SR}(n) = [I - \exp[n(M - I)]] \bar{e}_{SR}, \tag{3.8}$$

where

$$\bar{e}_{SR} \equiv [I - M]^{-1} m_2 \bar{\imath}_{SR}. \tag{3.9}$$

The reflective equilibrium outcomes for output, inflation and the nominal interest rate then converge as well as T is made large, to the values obtained by substituting the beliefs $e_{SR}(n)$ into the temporary equilibrium relations (2.12) and the reaction function (2.9).

The proof is given in section D of the Appendix. This result implies that our concept of reflective equilibrium, for any given degree of reflection n, has the intuitively appealing property that a commitment to follow a given policy for a time horizon T has similar consequences for all large enough values of T; moreover, for any large enough value of T, the policy that is expected to be followed after date T has little effect on equilibrium outcomes. Moreover,in the case of policies in the class considered here, there is no relevant difference between a commitment to a given reaction function for a long but finite time and a commitment to follow the rule forever.

Next, we consider how the reflective equilibrium prediction in the case of a long horizon T changes as the degree of reflection n increases. For the same reason that (3.7) holds, we can show that $e_{SR}(n) \to \bar{e}_S R$ as n is made large. Thus we obtain the following. (See section B and D of the Appendix for details.)

Proposition 3 *Suppose that in addition to the hypotheses of Proposition 2, the coefficients* (ϕ_π, ϕ_y) *satisfy the Taylor Principle (3.1). Then the limits*

$$\lim_{n\to\infty} \lim_{T\to\infty} e_t(n) \;=\; \lim_{n\to\infty} e_{SR}(n) \;=\; \bar{e}_{SR}$$

and

$$\lim_{T\to\infty} \lim_{n\to\infty} e_t(n) \;=\; \lim_{T\to\infty} e_t^{PF} \;=\; \bar{e}_{SR}$$

are well-defined and equal to one another. Moreover, both are independent of t*, and equal to the FS-PFE expectations in the case of a permanent commitment to the reaction function (2.9) with* $\bar{\imath}_t = \bar{\imath}_{SR}$*.*

Proposition 3 identifies a case in which perfect foresight analysis of the implications of a permanent commitment to a given policy rule does not lead to paradoxical conclusions. Not only does the question have a unique, well-behaved answer, which is the FS-PFE solution, but this answer provides a good approximation to the reflective equilibrium outcome in the case of any large enough degree of reflection n *and* any long enough horizon T for maintenance of the policy. As we shall see, however, the situation is different in the case of a commitment to a fixed nominal interest rate.

4 Consequences of a Temporarily Fixed Nominal Interest Rate

We now consider the case in which it comes to be understood that the nominal interest rate will be fixed at some level $\bar{\imath}_{SR}$ up to some date T, while it will again be determined by the "normal" central bank reaction function from date T onward — by which we mean a rule of the form (2.9), in which the response coefficients satisfy the Taylor Principle (3.1), and the intercept is consistent with the inflation target π^*. There are various reasons for interest in this case. First, a real disturbance may create a situation in which the interest rate prescribed by the Taylor rule violates the ZLB; in such a case, it may be reasonable to suppose that the central bank will set the nominal interest rate at the lowest possible rate as long as the situation persists, but return to implementation of its normal reaction function once this is feasible. And second, a central bank may commit itself to maintain the nominal interest rate at

its lower bound for a specific period of time, even if this is lower than the rate that the Taylor rule would prescribe; this was arguably the intention of the "date-based forward guidance" provided by the Fed and other central banks.

We are interested in two kinds of questions about the effects of such policies. One is what the effect should be of changing $\bar{\imath}_{SR}$, taking the horizon T as given. While there might seem to be no room to vary the short-run level of the interest rate, if we imagine a case in which it is already at the ZLB, it would even in that case always be possible to commit to a *higher* (though still fixed) interest-rate target, and some have suggested that inflation could be increased by doing so. A second question is the effect of changing the length of time that the interest rate is held fixed. To what extent can a commitment to keep the interest rate low for a longer time substitute for an ability to cut rates more sharply right away?

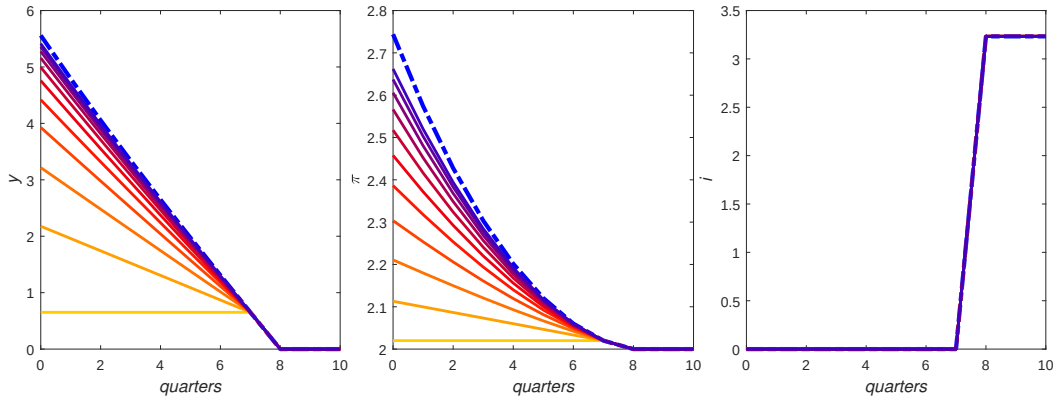
4.1 Convergence to Perfect-Foresight Equilibrium

We first consider whether reflective equilibrium converges to a perfect foresight equilibrium as n grows, and if so to which of the possible PFE paths. Because of the forward-looking character of our model and since we again assume a reaction function that satisfies the Taylor Principle from period T onward, the results of section 3 continue to apply. Specifically, Proposition 1 implies that in the case of any initial conjecture that converges exponentially, reflective equilibrium outcomes will converge to the unique FS-PFE outcomes as n increases. Note that this result already tells us that if the reflective equilibrium converges to *any* perfect foresight equilibrium, it can only converge to the FS-PFE.

The analysis of convergence prior to T, however, requires an extension of our previous result, because now the response coefficients (ϕ_π, ϕ_y) differ before and after T. Nonetheless, as shown in section D of the Appendix, the methods used to prove Proposition 1 can be extended to establish convergence in this case as well.

Proposition 4 *Consider the case of a shock sequence $\{g_t\}$ that converges exponentially, and let the forward path of policy be specified by a fixed interest rate $\bar{\imath}_{SR}$ for all $0 \leq t < T$, but by a reaction function of the form (2.9) for all $t \geq T$, where the coefficients (ϕ_π, ϕ_x) of the latter function satisfy (3.1), and the intercept is consistent with the inflation target π^*. Then in the case of any initial conjecture $\{e_t(0)\}$ regarding*

Figure 3: Reflective equilibrium for short term fixed interest rate



Note: The graph shows reflective equilibrium outcomes for $n = 0$ through 4 (progressively darker lines) compared with the FS-PFE solution (dash-dotted line), when the nominal interest rate is fixed for 8 quarters. See section F of the Appendix for details.

average expectations that converges exponentially, the belief revision dynamics (2.13) converge as n grows without bound to the belief sequence $\{e_t^{PF}\}$ *associated with the FS-PFE.*

The implied reflective equilibrium paths for output, inflation and the nominal interest rate similarly converge to the FS-PFE paths for these variables. This means that for any $\epsilon > 0$*, there exists a finite* $n(\epsilon)$ *such that for any degree of reflection* $n > n(\epsilon)$*, the reflective equilibrium value will be within a distance* ϵ *of the FS-PFE prediction for each of the three variables and at all horizons* $t \geq 0$*.*

Figure 3 provides a numerical illustration of this result. Again we assume the same parameters, $g_t = 0$ for all t and $e_t(0) = 0$ for all t. It is assumed that monetary policy will to depart from the "normal" Taylor rule for 8 quarters, and then to revert to the "normal" reaction function thereafter. The only difference is that in Figure 3 it is assumed that the nominal interest rate is fixed at zero for the first 8 quarters.

For the case $n = 0$ (the lightest of the lines in the figure), the responses are identical to those in Figure 2: both shifts in monetary policy have been chosen to lower the nominal interest rate to zero in the absence of any change in average expectations. For higher values of n, the effects of the policy change are qualitatively similar to those in Figure 2, but the output and inflation increases are somewhat larger when the interest rate is expected to remain fixed. This is because there is no endogenous interest-rate

increases in response to higher output and inflation.

As in the exercises of the previous section, the effects are larger the greater the degree of reflection and the longer the time for which the interest rate is expected to remain fixed, and are strongest under perfect foresight. However, the difference between the perfect foresight predictions and those from a given finite degree of reflection is greater than in the case of a temporary shift in the Taylor-rule intercept.

In Figure 3, as in Figure 2, an average degree of reflection of $n = 4$ results in temporary equilibrium outcomes that are similar to the perfect foresight predictions. But the reflective equilibrium outcomes when $n = 2$ are not as close to the perfect foresight outcomes as they are in Figure 2, especially in the first quarters. In quarter zero, the output response when $n = 2$ is 14 percent smaller than the perfect foresight prediction, and the inflation response is 10 percent smaller; and even when $n = 4$, the output and inflation responses are both about 3 percent smaller than the perfect foresight predictions. Moreover, these discrepancies rapidly become much larger if the interest rate is expected to be fixed for longer.

4.2 Very Long Periods with a Fixed Nominal Interest Rate

In the case of a temporary commitment to a lower Taylor-rule intercept, Proposition 4 implies that there exists a finite level of reflection n for which the reflective equilibrium predictions and the perfect foresight predictions are similar, regardless of how long the alternative policy is expected to last. This is no longer true in the case of a temporary commitment to a fixed low interest rate. In this case, we find that regardless of how high the (finite) level of reflection n may be, the reflective equilibrium predictions and the perfect foresight predictions are very different if T is large enough.

In particular, as shown above, the FS-PFE solution of our linearized model implies that the effects on both output and inflation grow without bound as the horizon T is extended farther into the future. Instead, in the case of any finite n, the effects on both variables predicted by the reflective equilibrium remain bounded. Methods similar to those used to establish Proposition 2 also allow us to show the following.

Proposition 5 *Consider the case in which $g_t = 0$ for all t, and let the forward path of policy be specified as in Proposition 4. Then if the initial conjecture is given by $e_t(0) = 0$ for all t, the reflective equilibrium beliefs $\{e_t(n)\}$ for any degree of reflection n converge*

to a well-defined limiting value

$$e_{SR}(n) \equiv \lim_{T\to\infty} e_t(n)$$

that is independent of t, and this limit is again given by (3.8), where $\bar{e}_{SR}$ is again defined in (3.9). The reflective equilibrium outcomes for output, inflation and the nominal interest rate then converge as well as T is made large, to the values obtained by substituting the beliefs $e_{SR}(n)$ into the temporary equilibrium relations (2.12) and the reaction function (2.9).

The proof is given in section D of the Appendix. The result is similar to the one stated in Proposition 2. It should be recalled that Proposition 2 does not require that the reaction function coefficients satisfy (3.1); it would apply, in particular, to the case $\phi_\pi = \phi_y = 0$, corresponding to fixed interest rates before and after date T. Proposition 5 establishes a similar result even when the response coefficients prior to date T differ from those from date T onward.

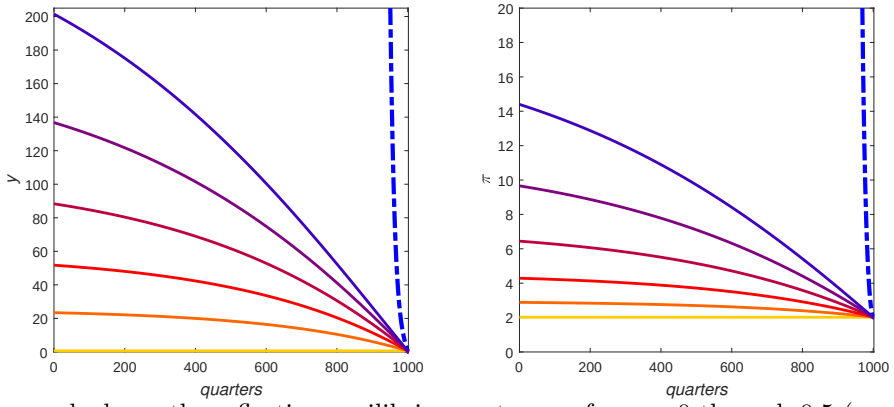
Thus if we consider the reflective equilibrium associated with any given finite n, the predicted outcomes are essentially the same for any long enough horizon T. There is no material difference between commitment to a fixed interest rate for a long but finite time and a permanent commitment. Hence our reflective equilibrium solution is not subject to the "failure of continuity at infinity" that seemed paradoxical in the case of the perfect-foreight analysis.

But this means that the perfect foresight predictions become more and more unlike the temporary equilibrium predictions (for any fixed n) when T is large. Figure 4 illustrates this, in the case of the same model calibration as used in previous figures, by considering a (certainly unrealistic) situation in which the nominal interest rate is expected to be fixed for 500 years.[38] This very long horizon is considered in order to make the backward convergence of the reflective equilibrium predictions evident. At such a horizon, the perfect-foresight predictions involve effects on output and inflation that are so much larger that they cannot be shown in the figure.[39]

[38] The third panel is omitted, since the path of the nominal interest rate is independent of the degree of reflection.

[39] There would be little point in even discussing the numerical values that are not shown, as our log-linearized equations cannot be expected to be accurate in the case of such gigantic departures from the point around which they have been log-linearized. We show the dash-dotted line to show

Figure 4: Reflective equilibrium when the interest rate is fixed for a long time.



Note: The graph shows the reflective equilibrium outcomes for $n = 0$ through 0.5 (progressively darker lines) compared with the FS-PFE solution (dash-dotted line), when the nominal interest rate is fixed for 250 years. See section F of the Appendix for details.

The reflective equilibrium prediction for a long commitment horizon T is not only quite different from the FS-PFE prediction for that horizon; it is *also* quite different from any of the perfect-foresight paths in the case of a permanent interest-rate peg. In particular, regardless of how long the duration of the commitment may be, and regardless of the degree of reflection, a commitment to keep the nominal interest rate at a low level for longer is both expansionary and inflationary.

Proposition 6 *For a given shock sequence* $\{g_t\}$ *and a given initial conjecture* $\{e_t(0)\}$*, consider monetary policies of the kind described in Proposition 4, with* $\bar{\imath}_{SR} < 0$ *(that is, an initial fixed interest rate at a level lower than the steady-state nominal interest rate associated with the long-run inflation target* π^**). Suppose also that* $g_t = 0$ *and* $e_t(0) = 0$ *for all* $t \geq T$.[40] *Then for any fixed* $\bar{\imath}_{SR}$ *and fixed level of reflection* $n > 0$*, increasing the length of the commitment from* T *to* $T' > T$ *increases both inflation and output in the reflective equilibrium, in all periods* $0 \leq t < T'$*, while it has no effect on either variable from date* T' *onward.*

The proof is given in section C of the Appendix. The qualitative prediction of the FS-

that the FS-PFE predictions are very different from the reflective equilibrium predictions.

[40] In fact, it should be evident from the proof given in section C of the Appendix that it suffices that $g_t \geq 0$ and $e_t(0) \geq 0$ for all $t \geq T$. What matters for the proof is that there not be factors tending to reduce output or inflation, apart from the effects of monetary policy, that are anticipated to affect periods beyond date T.

PFE analysis is confirmed — that is, the *signs* of the predicted effects on output and inflation are the same as in the FS-PFE analysis — even if the quantitative magnitude of the effects is quite different, especially when n is low.

It is important to note however that with a reflective equilibrium, when there is a negative shock to the economy, even committing to maintain a low interest rate *forever* may not suffice to prevent the economy from entering a recession. While the FS-PFE analysis implies that the effects of *any* size of contractionary shock can be completely counter-acted by a sufficiently long-lasting commitment to a low interest rate — and in fact, that a sufficiently long-lasting commitment can produce an inflationary boom of arbitrary size — it is possible, under the reflective equilibrium analysis, to find (if the degree of reflection is small enough) that even a promise to keep the interest rate *permanently* at zero would be insufficient to prevent output and inflation from both falling below their target values.[41]

4.3 The Source of the Paradoxes

We have seen that if we model the consequences of date-based forward guidance using our model of reflective equilibrium with some finite degree of reflection n, the paradoxical conclusions from perfect foresight equilibrium analysis are avoided. The predictions remain well-behaved as the length of duration of the commitment to the fixed interest rate is made arbitrarily long. The predictions in the case of all sufficiently long-duration commitments are essentially the same as one another, and the same as the prediction in the case of a permanent interest-rate peg. There are no material differences in the short-run outcomes predicted under policy commitments that differ only in what they prescribe for dates very far in the future. Nor is there any ambiguity about whether a commitment to fix interest rates at a lower-than-normal level for a longer period of time should be inflationary or deflationary. While the precise quantitative effects predicted for a given duration of commitment depend on the degree of reflection, a commitment to keep interest rates lower for longer will always be more inflationary.

The paradoxical conclusions in the case of perfect foresight analysis result from using the perfect foresight equilibrium solution concept in cases in which it is very

[41] See García-Schmidt and Woodford (2015, sec.4.4).

far from providing an accurate approximation to a temporary equilibrium with reflective expectations, even supposing that the degree of reflection is high. Under the assumption of a *permanent* interest-rate peg, the only forward-stable perfect foresight equilibria are ones that converge asymptotically to an inflation rate determined by the Fisher equation and the interest-rate target — and thus, that is lower by one percent for every one percent reduction in the interest rate. But for initial conjectures of the kind that we discuss above, *none* of these perfect foresight equilibria correspond, even approximately, to reflective equilibria — even to reflective equilibria for some very high degree of reflection n. Nor is this because in such cases high-n reflective equilibria correspond to some *other* kind of perfect foresight equilibrium; instead, one generally finds that the belief-revision dynamics fail to converge to *any* perfect foresight equilibrium as n increases, in the case of a permanent interest-rate peg.

This failure of convergence can be illustrated using results already presented above. Consider the case of a policy under which $i_t = \bar{\imath}_{LR}$ forever, and let us further assume that $g_t = 0$ for all t, and start from an initial conjecture under which $e_t = 0$ for all t. Then the belief-revision dynamics are given by $e_t(n) = e_{LR}(n)$ for all t, where $e_{LR}(n)$ evolves according to (2.19) starting from initial condition $e_{LR}(0) = 0$, and M in this equation is now the matrix corresponding to response coefficients $\phi_\pi = \phi_y = 0$.[42]

However, whereas in the Taylor-rule case considered in section 3, this solution implied that $e_{LR}(n) \to \bar{e}_{LR}$ as $n \to \infty$, this is no longer true in the case of an interest-rate peg. When $\phi_\pi = \phi_y = 0$, we show in section C of the Appendix that the matrix $M - I$ has a positive real eigenvalue. This in turn means that the elements of the matrix $\exp[n(M - I)]$ grow explosively as n is made large, and $e_{LR}(n)$ diverges from $\bar{e}_{LR}$. Nor does $e_{LR}(n)$ approach any perfect foresight equilibrium: the distance between $e_{LR}(n)$ and $e^*_{LR}(n)$ also grows explosively as n increases.

It similarly follows (using Proposition 5) that the nearly-stationary outcomes obtained in the case of any long enough finite-length interest-rate peg under a fixed degree of reflection n do not converge to any limit as n is made large. Thus neither of the double limits

$$\lim_{n\to\infty} \lim_{T\to\infty} e_t(n) \;=\; \lim_{n\to\infty} e_{SR}(n)$$

[42] The case considered is of the same kind as in section 3.3, except that we now set $T = 0$, and assume that $\phi_\pi = \phi_y = 0$.

or

$$\lim_{T\to\infty} \lim_{n\to\infty} e_t(n) \;=\; \lim_{T\to\infty} e_t^{PF}$$

exists in the case of a temporary interest-rate peg. (Both sequences diverge.) It is true that $\bar{e}_{SR}$ is still well-defined in this case. But $e_{SR}(n)$ no longer converges to it as n is made large, nor does e_t^{PF} as T is made large. Failure of the "Taylor Principle" invalidates *both* of those convergence results, relied upon in Proposition 3.

It remains true, for any finite length of peg, that a high enough degree of reflection leads to an outcome indistinguishable from the FS-PFE; and it is also true, for any finite degree of reflection, that a long enough finite-length peg leads to reflective equilibrium outcomes that are indistinguishable from those under a permanent peg. But it does *not* follow from these observations that a long enough peg together with a high enough degree of reflection must lead to anything similar to a forward-stable perfect foresight equilibrium associated with a permanent interest-rate peg. It is the failure to recognize this that leads to paradoxical conclusions in the argument sketched in the introduction.

5 Conclusion

Is there, then, reason to fear that a commitment to keep nominal interest rates low for a longer period of time will be deflationary, rather than inflationary? There is one way in which such an outcome could easily occur, and that is if the announcement of the policy change were taken to reveal negative information (previously known only to the central bank) about the outlook for economic fundamentals, rather than representing a pure change in policy intentions of the kind analyzed above.[43] This may well have been a problem with the way in which "date-based forward guidance" was used by the U.S. Federal Reserve during the period 2011-12, as discussed by Woodford (2013b); but it is not an inherent problem with announcing a change in future policy intentions.

We show that a commitment to keep nominal interest rates low for a longer time should be inflationary. If people believe the central bank's statements about its future policy intentions, and believe that it will indeed succeed in maintaining a low nominal

[43]For further discussion of the way in which the revelation of central-bank information by announced policy decisions can result in perverse effects, see García-Schmidt (2015).

interest rate, it does not follow that they must expect a deflationary equilibrium. This is so even if we suppose that they reason using a correct model of inflation and aggregate output determination.

If their reasoning occurs through a process of reflection of the kind modeled in this paper, then an increase in the expected length of time for which the nominal interest rate is expected to remain at some effective lower bound should result in expectations of higher income and higher inflation, regardless of the degree of reflection (as long as $n > 0$); and according to our model of temporary equilibrium resulting from optimizing spending and pricing decisions, such a change in expectations should result in higher output and inflation. This outcome may or may not approximate a perfect foresight equilibrium, depending on the degree of reflection; in the case of a commitment to keep the nominal interest rate low for a long enough period, it almost certainly will *not* resemble any perfect foresight equilibrium, even approximately.

Of course, we have here only provided an analysis of the short-run effects of such a policy announcement, before the expectations that it would make sense to assume for "naive" agents should change in response to new experience following the announcement. While an analysis of longer-run dynamics taking into account updating of the naive conjecture is beyond the scope of this paper, there is no reason to think that updating through an adaptive learning algorithm should eventually move the economy's evolution closer to the perfect foresight predictions. In the short run, we have shown, forward guidance should result in higher output and inflation than would otherwise occur; so if the naive conjecture is updated in the light of new experience, one would expect somewhat higher anticipated paths for output and inflation on the part of "naive" agents than if no forward guidance had been given. Such a change in the initial conjecture would only lead to even greater increases in output and inflation due to the forward guidance than the ones calculated above.

Thus we believe that it is important to explicitly model the process of belief revision as a result of further reflection, rather than assuming that a perfect foresight equilibrium must yield a correct prediction. Some macroeconomists may find the proposed alternative solution concept unappealing, since its prediction depends both on the initial conjecture and on how far the belief-revision process is followed. While this is true with regard to exact quantitative prediction, our approach gives unambiguous signs for the expected effects. Hence it is possible to obtain conclusions of a useful

degree of specificity, even when one has little ground for insisting on a single precise model of expectation formation.

It should also be noted that while our concept of reflective equilibrium can yield quite varied predictions under some circumstances, because the belief-revision dynamics diverge (or converge quite slowly), under other circumstances much tighter predictions are obtained, because of relatively rapid convergence of the belief-revision dynamics. It can then be a goal to choose a policy under which beliefs converge reliably, leading to less uncertainty about the outcome that should be expected.

In the case of a central bank that needs additional demand stimulus and is at the ZLB, announcing an intention to keep the interest rate at its lower bound for a long time, regardless of how economic conditions develop, is *not* an ideal policy response, according to this criterion. Such a policy should be expected to be stimulative, but the exact degree of stimulus is difficult to predict. It may not be possible to choose a length of time for which to commit that does not run simultaneously the risk of being too short to be effective, if n is too low, and wildly inflationary, if n is too high.

But one could achieve a less uncertain outcome, according to the reflective equilibrium analysis, by committing to maintain a low nominal interest rate until some macroeconomic target is reached, such as the price-level target proposed by Eggertsson and Woodford (2003).[44] In the case that people carry the belief-revision process forward to a high degree, they should expect interest rates to be raised relatively soon, but if instead they truncate the process at a relatively low degree of reflection, they should expected interest rates to remain low for much longer. In either case, belief that the central bank is serious about the policy should change expectations in a way that results in a substantial, but not extravagant, increase in current aggregate demand.

Thus even though the approach proposed here leads to a *set* of possible predictions in the case of a given policy specification rather than a *point* prediction, it still yields conclusions that are useful for policy design. Insisting on the use of perfect foresight analysis simply because it yields a more precise prediction can lead to large errors. One is reminded of the dictum of the British logician Carveth Read:[45] "It is better to be vaguely right than exactly wrong."

[44] This alternative to date-based forward guidance is also less likely to be misunderstood as revealing negative central-bank information, as discussed by Woodford (2013b).

[45] Read (1920, p. 351). The aphorism is often mis-attributed to John Maynard Keynes.

References

Andrade, Philippe, Gaetano Gaballo, Eric Mengus, and Benoît Mojon (2016). "Forward Guidance and Heterogeneous Beliefs". HEC Paris Research Paper no. ECO/SCD-2017-1192, November.

Angeletos, Marios and Chen Lian (2017). "Forward Guidance without Common Knowledge". NBER Working Paper no.22785, revised November.

Arad, Ayala and Ariel Rubinstein (2012). "The 11-20 Money Request Game: A Level-k Reasoning Study". *American Economic Review* 102 (7), pp. 3561–73.

Bullard, James (2010). "Seven Faces of 'The Peril'". *Federal Reserve Bank of St. Louis Review* 92 (5), pp. 339–352.

Camerer, Colin F., Teck-Hua Ho, and Juin-Kuan Chong (2004). "A Cognitive Hierarchy Model of Games". *The Quarterly Journal of Economics* 119 (3), pp. 861–898.

Cochrane, John H. (2011). "Determinacy and Identification with Taylor Rules". *Journal of Political Economy* 119 (3), pp. 565–615.

— (2017). "The New-Keynesian Liquidity Trap". unpublished, revised August.

Del Negro, Marco, Marc Giannoni, and Christina Patterson (2015). "The Forward Guidance Puzzle". Federal Reserve Bank of New York Staff Report no. 574, revised December.

Denes, Matthew, Gauti B. Eggertsson, and Sophia Gilbukh (2013). "Deficits, Public Debt Dynamics and Tax and Spending Multipliers". *The Economic Journal* 123 (566), F133–F163.

Eggertsson, Gauti B. and Michael Woodford (2003). "The Zero Bound on Interest Rates and Optimal Monetary Policy". *Brookings Papers on Economic Activity* 34 (1), pp. 139–211.

Evans, George W. and Bruce McGough (2017). "Interest Rate Pegs in New Keynesian Models". unpublished, March.

Evans, George W. and Garey Ramey (1992). "Expectation Calculation and Macroeconomic Dynamics". *The American Economic Review* 82 (1), pp. 207–224.

— (1995). "Expectation Calculation, Hyperinflation and Currency Collapse". *The New Macroeconomics: Imperfect Markets and Policy Effectiveness*. Ed. by H. D. Dixon and N. Rankin. Cambridge University Press, pp. 307–336.

— (1998). "Calculation, Adaptation and Rational Expectations". *Macroeconomic Dynamics* 2 (02), pp. 156–182.

Farhi, Emmanuel and Ivan Werning (2017). "Monetary Policy, Bounded Rationality, and Incomplete Markets". unpublished, September.

Galí, Jordi (2015). *Monetary Policy, Inflation and the Business Cycle, 2d ed.* Princeton University Press.

García-Schmidt, Mariana (2015). "Monetary Policy Surprises and Expectations". unpublished, Columbia University, August.

García-Schmidt, Mariana and Michael Woodford (2015). "Are Low Interest Rates Deflationary? A Paradox of Perfect-Foresight Analysis". NBER Working Paper no. 21614, October.

Guesnerie, Roger (1992). "An Exploration of the Eductive Justifications of the Rational-Expectations Hypothesis". *The American Economic Review* 82 (5), pp. 1254–1278.

— (2008). "Macroeconomic and Monetary Policies from the Eductive Viewpoint". *Monetary Policy Under Uncertainty and Learning*. Ed. by K. Schmidt-Hebbel and C. Walsh. Central Bank of Chile, pp. 171–202.

Hicks, John R. (1939). *Value and Capital*. Oxford: Clarendon.

Iovino, Luigi and Dmitriy Sergeyev (2017). "Central Bank Balance Sheet Policies without Rational Expectations". unpublished, Bocconi University, December.

Levin, Andrew, David López-Salido, Edward Nelson, and Tack Yun (2010). "Limitations on the Effectiveness of Forward Guidance at the Zero Lower Bound". *International Journal of Central Banking* 6 (1), pp. 143–189.

McKay, Alisdair, Emi Nakamura, and Jón Steinsson (2016). "The Power of Forward Guidance Revisited".

Nagel, Rosemarie (1995). "Unraveling in Guessing Games: An Experimental Study". *American Economic Review* 85 (5), pp. 1313–1326.

Preston, Bruce (2005). "Learning About Monetary Policy Rules when Long-Horizon Expectations Matter". *International Journal of Central Banking* 1 (1), pp. 81–126.

Read, Carveth (1920). *Logic: Deductive and Inductive*. Simpkin.

Schmitt-Grohé, Stephanie and Martín Uribe (2010). "Liquidity Traps: An Interest-Rate-Based Exit Strategy". NBER Working Paper no. 16514, November.

Taylor, John B. (1993). "Discretion versus Policy Rules in Practice". *Carnegie-Rochester Conference Series on Public Policy* 39 (1), pp. 195–214.

Werning, Ivan (2012). "Managing a Liquidity Trap: Monetary and Fiscal Policy". Unpublished, April.

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

— (2013a). "Macroeconomic Analysis Without the Rational Expectations Hypothesis". *Annual Review of Economics* 5 (1), pp. 303–346.

— (2013b). "Forward Guidance by Inflation-Targeting Central Banks". *Sveriges Riksbank Economic Review* 2013 (3), pp. 81–120.

Yun, Tack (1996). "Nominal Price Rigidity, Money Supply Endogeneity, and Business Cycles". *Journal of Monetary Economics* 37 (2), pp. 345 –370.

FOR ONLINE PUBLICATION

Appendix to García-Schmidt and Woodford, "Are Low Interest Rates Deflationary? A Paradox of Perfect-Foresight Analysis"

A Summary of Notation Used in the Paper

Var./Param.	Explanation
One dimensional parameters	
n	degree opf reflection, introduced in equation (2.3).
π^*	inlfation target, introduced in equation (2.4).
β	discount factor when the rate of time preference is $\bar{\rho}$, introduced in equation (2.4).
σ	Household intertemporal elasticity of substitution, introduced in equation (2.4).
α	firm's j probability of not optimizing its price, introduced in equation (2.7).
ξ	elasticity of a firm's optimal relative price with respect to aggregate demand, introduced in equation (2.7).
κ	parameter of aggregate supply, introduced in equation (2.7) and defined below.
ϕ_π	Taylor rule coefficient for inflation, introduced in equation (2.9).
ϕ_y	Taylor rule coefficient for output, introduced in equation (2.9).
δ_i	Definition of parameters for $i = \{1, 2\}$ introduced in equation (2.11) and defined below.
ϕ	Taylor rule coefficient for inflation in Simple Illustration, introduced in equation (2.16).
η_y	Coefficient of output expectation in static relation defined in equation (2.17).
η_π	Coefficient of inflation expectation in static relation defined in equation (2.17).
η_i	Coefficient of interest rate expectation in static relation defined in equation (2.17).
z_∞	Finite collection of real coefficients defined in equation (3.5).
u_k	Finite collection of real coefficients defined in equation (3.5).
λ_k	Real numbers defined in equation (3.5).

Var./Param.	Explanation
Matrix/vector parameters	
c	2x2 Matrix relating the exogenous vector ω_t to the endogenous vector x_t, introduced in (2.10).
C	2x2 Matrix relating the expectation vector e_t to the endogenous vector x_t, introduced in (2.10).
m	2x2 Matrix relating he exogenous vector ω_t to the vector a_t, introduced in (2.12).
M	2x2 Matrix relating he expectation vector e_t to the vector a_t, introduced in (2.12).
ζ_j	2x2 Matrix j of the FS-PFE, introduced in equation (3.4).
m_2	2x1 Matrix, that is the second column of m, introduced in (3.6).
Variables	
c_t^i	consumption of household i, introduced in equation (2.4).
$\hat{b}_t^i$	net real financial wealth of household i, introduced in equation (2.4).
y_t	output, introduced in equation (2.4).
i_t	interest rate, introduced in equation (2.4).
π_t	inflation, introduced in equation (2.4).
ρ_t	household's rate of time preference, introduced in equation (2.4).
g_t	Weighted sum of household's rate of time preference, introduced in equation (2.5) and defined below.
v_t^i	Expectational variable of household i, introduced in equation (2.5) and defined below.
e_{1t}	Average expectation of v_{t+1}^i, introduced in equation (2.6) and defined below.
p_t^{*j}	Optimal price in t of firm j in excess of the average prices that are not reconsidered, introduced in equation (2.7).
p_t	Price level in t, introduced in equation (2.7).
e_{2t}	Average expectation of p_{t+1}^j, introduced in equation (2.8) and defined below.
$\bar{\imath}_t$	Intercept Taylor rule, introduced in (2.9).
e_{it}^*	Correct value for subjective expectation e_{it} for $i = \{1, 2\}$, defined in equation (2.11).
a_{it}	Variable to calculate e_{it}^* for $i = \{1, 2\}$, introduced in equation (2.11) and defined below.
π^e	Expectation of future inflation in Simple Illustration, introduced in equation (2.15).

Var./Param.	Explanation
Variables	
y^e	Expectation of future output in Simple Illustration, introduced in equation (2.15).
i^e	Expectation of future interest rate in Simple Illustration, introduced in equation (2.15).
η	Expectational variable, introduced in equation (2.17).
z_t	General variable defined in equation (3.5).
Vectors	
x_t	Vector containing endogenous variables. Defined generically in equation (2.1) and as the vector containing y_t and π_t in equation (2.10).
$\mathbf{e}_t$	Infinite-dimensional vector of average expectations defined in equation (2.1)
$\mathbf{e^*}_t$	Correct values of $\mathbf{e}$ implied by equilibrium dynamics given $\mathbf{e}$. Defined in (2.2)
$\dot{\mathbf{e}}$	Derivative of $\mathbf{e}_t$ with respect to n defined in (2.3)
e_t	Vector containing expectational variables e_{1t} and e_{2t}, introduced in equation (2.10).
ω_t	Vector containing exogenous variables g_t and $\bar{\imath}_t$, introduced in equation (2.10).
a_t	Vector containing variables a_{1t} and a_{2t}, introduced in equation (2.12).
$e_t(n)$	Same as e_t, making explicit that it depends on the degree of reflection n, introduced in equation (2.13).
$e_t^*(n)$	Vector containing e_{1t}^* and e_{2t}^*, making explicit that it depends on the degree of reflection n, introduced in equation (2.13).
$\dot{e}_t(n)$	Derivative of $e_t(n)$ with respect to n, introduced in equation (2.13).
$\bar{e}$	Perfect foresight equilibrium for expectational variable e, introduced in equation (2.19).
$\dot{\eta}$	Derivative of η with respect to n, introduced in equation (2.20).

B Mathematical Derivations

B.1 Derivation of equation (2.4)-(2.6)

The economy is made up of a continuum of identical infinite-lived households indexed by $i \in [0, 1]$. Each household maximizes its estimate of its discounted utility:

$$\hat{E}_t^i \sum_{T=t}^{\infty} \exp[-\sum_{s=t}^{T-1} \hat{\rho}_s][u(C_T^i) - v(H_T^i)]$$

C_t^i is a Dixit-Stiglitz aggregate of the households' purchases of differentiated consumer goods, H_t^i is hours worked by the household in t, $\hat{\rho}_t$ is a possibly time-varying discount rate. It is assumed that the households supply the hours of work demanded by firms at a wage fixed by a union that bargains on behalf of households. This implies that its non-financial income (sum of wage income and share of profits) are outside its control. It is further assumed that the hours supplied by each household and its shares of the firms' profits is distributed equally among the household. Then, we can write the budget constraint of the household as:

$$B_{T+1}^i = (1 + \tilde{i}_T) \left[B_T^i + W_t H_T^i + \int_{j=0}^{1} \Pi_T(j) dj - P_T C_T^i \right]$$

with B_t^i bond holdings by household i at t, $\tilde{i}_t$ the interest rate of the bond holdings, W_t the wage, $\Pi_t(j)$ profits of firm j, P_t the price of the consumption basket. The problem of each household can be solved with the lagrangian:

$$\begin{aligned} \mathcal{L} \;=\; & \hat{E}_t^i \sum_{T=t}^{\infty} \exp[-\sum_{s=t}^{T-1} \hat{\rho}_s] \Bigg\{ u(C_T^i) - v(H_T^i) + \\ & \lambda_T \left((1 + \tilde{i}_T) \left[B_T^i + W_t H_T^i + \int_{j=0}^{1} \Pi_T(j) dj - P_T C_T^i \right] - B_{T+1}^i \right) \Bigg\} \end{aligned}$$

The FOCs can be written as:

$$\begin{array}{ll} [C_t^i] & U'(C_t^i) - (1 + \tilde{i}_t) P_t \lambda_t = 0 \\ [B_{t+1}^i] & -\lambda_t + \exp\{-\hat{\rho}_t\} \hat{E}_t^i (1 + \tilde{i}_{t+1}) \lambda_{t+1} = 0 \end{array}$$

Which implies the Euler equation:

$$U'(C_t^i) = \exp\{-\hat{\rho}_t\} (1 + \tilde{i}_t) \hat{E}_t^i \frac{U'(C_{t+1}^i)}{\Pi_{t+1}}$$

with $\Pi_t = P_t / P_{t-1}$. By replacing the equations for the profits, using the market clearing of the labor market, and dividing by P_{t-1}, we get that the budget constraint can be written as:

$$b_{t+1}^i = (1 + \tilde{i}_t) \left[\frac{b_t^i}{\Pi_t} + Y_t - C_t^i \right]$$

where $Y_t = \int_{j=0}^{1} Y_t(j) dj$, $b_t^i = B_t^i / P_{t-1}$.

The steady state in which these equations will be log-linearized is one with positive

inflation, $\pi > 1$. The approximations are given by:

$$\hat{b}^i_{t+1} \approx \frac{1}{\beta}\hat{b}^i_t + \frac{\pi}{\beta}\left(y_t - c^i_t\right)$$

With $\hat{b}^i_t = b^i_t - b$, $y_t = \log(Y_t/Y)$, $c^i_t = \log(C^i_t/C)$ and all the variables without time subscript are steady state values. This equation uses the fact that in steady state $b^i = b = 0$, $Y = C$ and $(1 + \tilde{i}) = \pi/\beta$. This implies:

$$\hat{b}^i_t = -\pi \sum_{T=t}^{\infty} \beta^{T-t} \hat{E}^i_t (y_T - c^i_T)$$

The approximation of the Euler equation:

$$\hat{E}^i_t c^i_{t+1} = c^i_t + \sigma(i_t - \rho_t - \hat{E}^i_t \pi_{t+1})$$

with $\pi_t = \log(\Pi_t/\pi)$ and $\rho_t = \hat{\rho}_t - \hat{\rho}$, $i_t = \log(1 + \tilde{i}_t) - \log(1 + \tilde{i})$. This implies:

$$\hat{E}^i_t c^i_T = c^i_t + \sigma \sum_{s=t}^{T-1} \hat{E}^i_t (i_s - \rho_s - \pi_{s+1})$$

And writing everything together:

$$\begin{aligned}
\hat{b}^i_t &= -\pi \sum_{T=t}^{\infty} \beta^{T-t} \hat{E}^i_t y_T + \pi \left(\sum_{T=t}^{\infty} \beta^{T-t} \left(c^i_t + \sigma \sum_{s=t}^{T-1} \hat{E}^i_t (i_s - \rho_s - \pi_{s+1}) \right) \right) \\
&= -\pi \sum_{T=t}^{\infty} \beta^{T-t} \hat{E}^i_t y_T + \frac{\pi c^i_t}{1-\beta} + \frac{\sigma\pi\beta}{1-\beta} \sum_{T=t}^{\infty} \beta^{T-t} \hat{E}^i_t (i_T - \rho_T - \pi_{T+1}) \\
c^i_t &= \frac{1-\beta}{\pi} \hat{b}^i_t + \sum_{T=t}^{\infty} \beta^{T-t} \hat{E}^i_t \left((1-\beta) y_T - \beta\sigma (i_T - \rho_T - \pi_{T+1}) \right)
\end{aligned}$$

which is equation (2.4). Then the change to equation (2.5) is direct and also the aggregation to (2.6) by realizing that $\int \hat{b}^i_t di = 0$ and $\int c^i_t di = y_t$.

B.2 Derivation of equation (2.7)

Consider a firm j which uses labor to produce its product,

$$Y_t(j) = f(H_t(j))$$

where $Y_t(j)$ is firm j's product, $f(.)$ is its production technology and $H_t(j)$ is the labor used by the firm. Consider also that this firm faces a downward sloping demand because it produces a differentiated product:

$$Y_t(j) = Y_t \left(\frac{P_t(j)}{P_t} \right)^{-\theta}$$

where Y_t is aggregate demand, $P_t(j)$ is firm j's price. We can then write the profit of firm j as:

$$\Pi_t(j) = \Pi \left(P_t(j), P_t, Y_t, W_t \right)$$

The problem of choosing the price optimally has to take into account that prices, when not chosen are revised by the inflation target, so we can write the maximization objective as:

$$\max \hat{E}_t^j \sum_{T=t}^{\infty} \alpha^{T-t} Q_{t,T} \Pi \left(P_t(j) (\Pi^*)^{T-t}, P_T, Y_T, W_T \right)$$

where $Q_{t,T}$ is the household's stochastic discount factor. Using the homogeneity of degree zero in prices of the derivative of $\Pi(.)$ with respect to its first argument, $\Pi_1(.)$, the log-linearized version of the optimal condition of labor from the household and market clearing, we get the log-linearized FOC of this function[46]:

$$\hat{E}_t^j \sum_{T=t}^{\infty} (\alpha\beta)^{T-t} \left(\log P_t^*(j) + (T-t) \log \Pi^* - \log P_T - \xi \log Y_T / Y \right) = 0$$

This gives you equation (2.7) noting that $p_t = \log P_t$, $p_t^{*j} = \log P_t^j(j) - p_{t-1} - \pi^*$, $\pi^* = \log(\Pi^*)$.

B.3 Derivation of equation (2.8)

First note that the price index evolves according to:

$$\begin{aligned} p_t &= \int_{j=0}^{\alpha} (p_{t-1} + \pi^*) dj + \int_{j=\alpha}^{1} (p_t^{*j} + p_{t-1} + \pi^*) dj \\ p_t - p_{t-1} - \pi^* &= \int_{j=\alpha}^{1} p_t^{*j} dj \end{aligned}$$

so:

$$\pi_t = (1-\alpha) \int_{j=0}^{1} p_t^{*j} dj \tag{B.1}$$

[46] To see details of this derivation, refer to Woodford 2003, chap. 3.

since $\pi_t = p_t - p_{t-1} - \pi^*$. Starting from (2.7), we have:

$$\begin{aligned}
p_t^{*j} &= (1-\alpha\beta)\sum_{T=t}^{\infty}(\alpha\beta)^{T-t}\hat{E}_t^j[p_T + \xi y_T - \pi^*(T-t)] - (p_{t-1} + \pi^*) \\
&= (1-\alpha\beta)\sum_{T=t}^{\infty}(\alpha\beta)^{T-t}\hat{E}_t^j\xi y_T + (1-\alpha\beta)\sum_{T=t}^{\infty}(\alpha\beta)^{T-t}\hat{E}_t^j[p_T - \pi^*(T-t) - p_{t-1} - \pi^*] \\
&= (1-\alpha\beta)\sum_{T=t}^{\infty}(\alpha\beta)^{T-t}\hat{E}_t^j\xi y_T + (1-\alpha\beta)\hat{E}_t^j\Big[p_t - p_{t-1} - \pi^* + \alpha\beta(p_{t+1} - p_{t-1} - 2\pi^*) + \\
&\quad (\alpha\beta)^2(p_{t+2} - p_{t-1} - 3\pi^*)...\Big] \\
&= (1-\alpha\beta)\sum_{T=t}^{\infty}(\alpha\beta)^{T-t}\hat{E}_t^j\xi y_T + (1-\alpha\beta)\hat{E}_t^j\left[\pi_t + \alpha\beta(\pi_{t+1} + \pi_t) + (\alpha\beta)^2(\pi_{t+2} + \pi_{t+1} + \pi_t)...\right]
\end{aligned}$$

which we can write as:

$$p_t^{*j} = \sum_{T=t}^{\infty}(\alpha\beta)^{T-t}\hat{E}_t^j\left[\pi_T + (1-\alpha\beta)\xi y_T\right] \tag{B.2}$$

This implies:

$$p_t^{*j} = \pi_t + (1-\alpha\beta)\xi y_t + \alpha\beta\hat{E}_t^j p_{t+1}^{*j}$$

Integrating over firms we have:

$$\int_{j=0}^{1} p_t^{*j}dj = \frac{\pi_t}{1-\alpha} = \pi_t + (1-\alpha\beta)\xi y_t + \alpha\beta\int_{j=0}^{1}\hat{E}_t^j p_{t+1}^{*j}dj$$

Which by multiplying by $(1-\alpha)$, defining κ and rearranging terms gives you equation (2.8). Equation (2.11) is obtained directly by the definition of v_t^i in the text and (B.2).

B.4 Derivation of matrices and equation (2.10)

Replacing (2.9) in (2.6) and (2.8), we get the system:

$$C_1 x_t = C_2 e_t + C_3\omega_t$$

with

$$x_t = \begin{bmatrix} y_t \\ \pi_t \end{bmatrix} \qquad e_t = \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix} \qquad \omega_t = \begin{bmatrix} g_t \\ \bar{\imath}_t \end{bmatrix}$$

and

$$C_1 = \begin{bmatrix} 1+\sigma\phi_y & \sigma\phi_\pi \\ -\kappa & 1 \end{bmatrix} \qquad C_2 = \begin{bmatrix} 1 & 0 \\ 0 & (1-\alpha)\beta \end{bmatrix} \qquad C_3 = \begin{bmatrix} 1 & -\sigma \\ 0 & 0 \end{bmatrix}$$

Which, by inverting and pre-multiplying C_1, gives you (2.10), with the matrices:

$$C = \tfrac{1}{\Delta}\begin{bmatrix} 1 & -\sigma\phi_\pi(1-\alpha)\beta \\ \kappa & (1+\sigma\phi_y)(1-\alpha)\beta \end{bmatrix}, \qquad c = \tfrac{1}{\Delta}\begin{bmatrix} 1 & -\sigma \\ \kappa & -\kappa\sigma \end{bmatrix},$$

and use the shorthand notation $\Delta \equiv 1 + \sigma\phi_y + \sigma\kappa\phi_\pi \geq 1$. (This last inequality, that allows us to divide by Δ, holds under the sign restrictions maintained in the text.) Given this solution for x_t, the solution for the nominal interest rate is obtained by substituting the solutions for inflation and output into the reaction function (2.9). You can check that $C = C_1^{-1}C_2$, $c = C_1^{-1}C_3$.

B.5 Derivation of equation (2.11)

The definition of $e_{1,t}$ is given by:

$$\begin{aligned} e_{1t} &= \int \hat{E}_t^i v_{t+1}^i di \\ v_t^i &= \sum_{T=t}^{\infty} \beta^{T-t} \hat{E}_t^i \{(1-\beta)y_T - \sigma(\beta i_T - \pi_T)\} \end{aligned}$$

Lets call e_{1t}^*, the implied value of e_{1t} when we actually replace the values of $\{y_t, \pi_t, i_t\}$ that are calculated using beliefs $\{e_{1t}, e_{2t}\}$

$$\begin{aligned} e_{1t}^* &= \int \hat{E}_t^i \sum_{T=t+1}^{\infty} \beta^{T-t-1} \hat{E}_{t+1}^i \{(1-\beta)y_T - \sigma(\beta i_T - \pi_T)\} di \\ &= (1-\beta) \sum_{T=t+1}^{\infty} \beta^{T-t-1} \int \hat{E}_t^i \{y_T - \frac{\sigma}{1-\beta}(\beta i_T - \pi_T)\} di \\ &= (1-\delta_1) \sum_{T=t+1}^{\infty} \delta_1^{T-t-1} \bar{E}_t \left\{ y_T - \frac{\sigma}{1-\beta}(\beta i_T - \pi_T) \right\} \end{aligned}$$

where $\delta_1 = \beta$ and $\bar{E}_t$ is the average expectation. For the second expectational value we follow the same steps, given the definitions provided in the text, but using the

equation for the optimal price in (B.2), we have:

$$\begin{aligned} e_{2t} &= \int \hat{E}_t^j p_{t+1}^{*j} dj \\ p_t^{*j} &= \sum_{T=t}^{\infty} (\alpha\beta)^{T-t} \hat{E}_t^j \{\pi_T + (1-\alpha\beta)\xi y_T)\} \end{aligned}$$

we call e_{2t}^*, the implied value of e_{2t} when we actually replace the values of $\{y_t, \pi_t, i_t\}$ that are calculated using beliefs $\{e_{1t}, e_{2t}\}$

$$\begin{aligned} e_{2t} &= \int \hat{E}_t^j \sum_{T=t+1}^{\infty} (\alpha\beta)^{T-t-1} \hat{E}_{t+1}^j \{\pi_T + (1-\alpha\beta)\xi y_T)\} dj \\ &= (1-\alpha\beta) \sum_{T=t+1}^{\infty} (\alpha\beta)^{T-t-1} \int \hat{E}_t^j \left\{ \frac{1}{1-\alpha\beta} \pi_t + \xi y_T \right\} dj \\ &= (1-\delta_2) \sum_{T=t+1}^{\infty} (\alpha\beta)^{T-t-1} \bar{E}_t \left\{ \frac{1}{1-\alpha\beta} \pi_t + \xi y_T \right\} \end{aligned}$$

where $\delta_2 = \alpha\beta$ and $\bar{E}_t$ is the average expectation. It is assumed that the average expectation of households and firms are the same.

B.6 Derivation of equation (2.12): matrices m and M

Starting from the definitions of a_{1t} and a_{2t}, and replacing (2.9), we can write the system as:

$$a_t = M_1 x_t + m_1 \omega_t$$

with

$$M_1 = \begin{bmatrix} 1 - \frac{\beta\sigma\phi_y}{1-\beta} & \frac{\sigma}{1-\beta}(1-\sigma\phi_\pi) \\ \xi & \frac{1}{1-\alpha\beta} \end{bmatrix} \qquad m_1 = \begin{bmatrix} 0 & -\frac{\beta\sigma}{1-\beta} \\ 0 & 0 \end{bmatrix}$$

we can replace x_t by (2.10) to get (2.12) with:

$$M = \frac{1}{\Delta} \begin{bmatrix} \frac{1+\sigma\kappa-\beta\Delta}{1-\beta} & \frac{\sigma\beta(1-\alpha)(1+\sigma\phi_y-\phi_\pi)}{1-\beta} \\ \frac{\kappa}{(1-\alpha)(1-\alpha\beta)} & \frac{\beta(1+\sigma\phi_y-\alpha\Delta)}{1-\alpha\beta} \end{bmatrix}, \qquad m = \frac{1}{\Delta} \begin{bmatrix} \frac{1+\sigma\kappa-\beta\Delta}{1-\beta} & -\frac{\sigma(1+\sigma\kappa)}{1-\beta} \\ \frac{\kappa}{(1-\alpha)(1-\alpha\beta)} & -\frac{\sigma\kappa}{(1-\alpha)(1-\alpha\beta)} \end{bmatrix}.$$

where you can check that $M = M_1 C$ and $m = M_1 c + m_1$.

Putting together the equations (2.11) and (2.12) we can write the equation for e_t^*

as follows:

$$e_t^* = (I - \Lambda) \sum_{j=1}^{\infty} \Lambda^{j-1} \left[M e_{t+j} + m \omega_{t+j} \right] \tag{B.3}$$

for all $t \geq 0$, where the sequences of matrices $\{\psi_j\}$ and $\{\varphi_j\}$ are given by

$$\Lambda \equiv \begin{pmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{pmatrix}$$

for all $j \geq 1$.

B.7 Derivations of the Simple Illustration

As given in the text, we have that the temporary equilibrium relations are given by:

$$\begin{aligned} y &= -\sigma i + e_1 \\ \pi &= \kappa y + (1 - \alpha)\beta e_2 \\ i &= \bar{\imath} + \phi \pi \end{aligned}$$

Where (e_1, e_2) are given by their definitions, which in this case becomes:

$$\begin{aligned} e_1 &= \int \sum_{T=t+1}^{\infty} \beta^{T-t-1} \hat{E}_t^i \{(1-\beta) y_T - \sigma(\beta i_T - \pi_T)\} di = \sum_{T=t+1}^{\infty} \beta^{T-t-1} \{(1-\beta) y^e - \sigma(\beta i^e - \pi^e)\} \\ &= y^e - \frac{\sigma}{1-\beta}(\beta i^e - \pi^e) \end{aligned}$$

and

$$\begin{aligned} e_2 &= \int \sum_{T=t+1}^{\infty} (\alpha\beta)^{T-t-1} \hat{E}_t^j \{\pi_T + (1-\alpha\beta)\xi y_T\} dj = \sum_{T=t+1}^{\infty} (\alpha\beta)^{T-t-1} \{\pi^e + (1-\alpha\beta)\xi y^e\} \\ &= \frac{\pi^e}{1-\alpha\beta} + \xi y^e \end{aligned}$$

To follow the notation given in the rest of the paper, lets replace the monetary policy in the other two equations to get:

$$C_1 x = C_2 e + c_3' \bar{\imath}$$

with

$$x \equiv \begin{bmatrix} y \\ \pi \end{bmatrix}, \qquad e = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

and C_1, C_2 are the same as before just replacing $\phi_y = 0$ and $\phi_\pi = \phi$ and c_3' is the second column of c_3. By inverting the first matrix, we get the equivalent to (2.10)

$$x = Ce + c_2\bar{\imath}$$

with c_2 the second column of matrix c and again the coefficients are replaced so that $\phi_y = 1$ and $\phi_\pi = \phi$. Given e, we have that the values of the endogenous variables is given by the above system and (2.9). To go to the next step and update the beliefs, we need the values for (e_1^*, e_2^*). Note in (2.11) that in this case $e_i^* = a_i$, since the expectation of all future variables is the same. By their definitions, we have then:

$$\begin{aligned} e_1^* &= y - \frac{\sigma}{1-\beta}(\beta i - \pi) \\ e_2^* &= \frac{\pi}{1-\alpha\beta} + \xi y \end{aligned}$$

Note that this is the same as the equations for (e_1, e_2) given previously, just replacing the actual values by the expected values. By replacing the equation for the interest rate, we can write this as:

$$e^* = M_1 x + m_1'\bar{\imath}$$

with

$$e^* = \begin{bmatrix} e_1^* \\ e_2^* \end{bmatrix}$$

where M_1 is the same as the one defined in a previous subsection of this Appendix, just replacing $\phi_y = 0$ and $\phi_\pi = \phi$ and m_1' is the second column of matrix m_1. By replacing the TE relations can be written as:

$$e^* = Me + m_2\bar{\imath}$$

with M, the same as before, just replacing $\phi_y = 0$ and $\phi_\pi = \phi$, and m_2 is the second column of m replacing the same parameters as in M. Replacing e^* in (2.13) gives you

$$\begin{aligned} \dot{e} &= e^* - e = Me + m_2\bar{\imath} - e \\ &= (M - I)(e - (I - M)^{-1} m_2\bar{\imath}) \end{aligned}$$

which becomes (2.19) since $\bar{e}$ is the solution of the previous equation by setting $\dot{e}$ to

zero:

$$\begin{aligned}(I-M)\bar{e} &= m_2\bar{\imath} \\ \bar{e} &= (I-M)^{-1}m_2\bar{\imath}\end{aligned}$$

which is the rest point of the previous system as long as $(I-M)$ is invertible.

The properties of the eigenvalues of $M-I$ are discussed in section C. It is shown that the real parts of $M-I$ are negative as long as the Taylor principle is satisfied, which in this case is $\phi > 1$. When the condition $\phi > 1$ is not met, one of the eigenvalues is positive.

B.8 Derivation of PFE equations

B.8.1 Neo-Keynesian IS curve: equation (3.2)

Starting from (2.4) we aggregate over households and we get:

$$\begin{aligned}y_t &= \sum_{T=t}^{\infty}\beta^{T-t}E_t\{(1-\beta)y_T - \beta\sigma(i_T - \pi_{T+1} - \rho_T)\} \\ &= (1-\beta)y_t - \beta\sigma(i_t - \pi_{t+1} - \rho_t) + \beta E_t y_{t+1}\end{aligned}$$

Which simplifying and rearranging gives you equation (3.2). To get to this same equation from (2.6) takes a little longer and need to rearrange more terms.

$$\begin{aligned}y_t &= \sigma\sum_{T=t}^{\infty}\beta^{T-t}\rho_T - \sigma i_t + \sum_{T=t+1}^{\infty}\beta^{T-t-1}E_t((1-\beta)y_T - \sigma(\beta i_T - \pi_T)) \\ y_t &= (1-\beta)y_t + \beta y_t \\ &= (1-\beta)y_t + \beta\left(\sigma\sum_{T=t}^{\infty}\beta^{T-t}\rho_T - \sigma i_t + \sum_{T=t+1}^{\infty}\beta^{T-(t+1)}E_t((1-\beta)y_T - \sigma(\beta i_T - \pi_T))\right) \\ &= (1-\beta)y_t - \beta\sigma(i_t - \rho_t - E_t\pi_{t+1}) + \beta\left(\sum_{T=t+1}^{\infty}\beta^{T-(t+1)}E_t\left[(1-\beta)y_T - \beta\sigma(i_T - \rho_T - \pi_{T+1})\right]\right) \\ &= (1-\beta)y_t - \beta\sigma(i_t - \rho_t - E_t\pi_{t+1}) + \beta E_t y_{t+1}\end{aligned}$$

And rearranging and dividing by β gives you equation (3.2).

B.8.2 Neo-Keynesian Phillips curve: equation (3.3)

First note that equation (B.1) is:

$$\pi_t = (1-\alpha)p_t^*$$

Since p_t^{*j} is the same for all j. Using this and replacing in equation (B.2), we get:

$$\frac{\pi_t}{1-\alpha} = \pi_t + (1-\alpha\beta)\xi y_t + \alpha\beta\frac{E_t\pi_{t+1}}{1-\alpha}$$

which, rearranging terms and defining κ gives you equation (3.3).

B.8.3 Derivation of the 2x2 system of the PFE and and equation (3.4)

By replacing equation (2.9) in equations (3.2) and (3.3), we get the system:

$$C_1 x_t = A_2 x_{t+1} + a(\rho_t - \bar{\imath}_t)$$

with

$$A_2 = \begin{bmatrix} 1 & \sigma \\ 0 & \beta \end{bmatrix} \qquad a = \begin{bmatrix} \sigma \\ 0 \end{bmatrix}$$

By inverting and pre-multiplying A_2, you can write this system as:

$$x_t \;=\; B\,x_{t+1} + b\,(\rho_t - \bar{\imath}_t) \tag{B.4}$$

where we define

$$B = \tfrac{1}{\Delta}\begin{bmatrix} 1 & \sigma(1-\beta\phi_\pi) \\ \kappa & \sigma\kappa + \beta(1+\sigma\phi_y) \end{bmatrix}, \qquad b = \tfrac{1}{\Delta}\begin{bmatrix} \sigma \\ \sigma\kappa \end{bmatrix}.$$

As shown in section C, when (3.1) is satisfied, this system has a unique bounded solution, since both eigenvalues of matrix B have modulus less than 1. This solution is given by (3.4) with

$$\zeta_j = B^j b$$

To obtain the same 2x2 system from the equations defining the Temporary equilibrium, we need to impose e_t must equal e_t^* for all t. From (B.3) it follows that a

sequence of vectors of expectations $\{e_t\}$ constitute PFE expectations if and only if

$$\begin{aligned} e_t &= e_t^* = (I-\Lambda)\sum_{j=1}^{\infty}\Lambda^{j-1}\left[Me_{t+j}+m\omega_{t+j}\right] \\ &= +(I-\Lambda)(Me_{t+1}+m\omega_{t+1});+\Lambda\, e_{t+1} \\ &= \left[(I-\Lambda)M\ +\ \Lambda\right]e_{t+1}\ +\ (I-\Lambda)m\,\omega_{t+1} \end{aligned} \tag{B.5}$$

for all $t \geq 0$.

The dynamics implied by (B.5) are in fact equivalent to those implied by (B.4). Using (2.10) together with (B.5) implies that the PFE dynamics of output and inflation must satisfy

$$\begin{aligned} x_t &= C\left[(I-\Lambda)M\ +\ \Lambda\right]e_{t+1}\ +\ C(I-\Lambda)m\,\omega_{t+1}\ +c\,\omega_t \\ &= C\left[(I-\Lambda)M\ +\ \Lambda\right]C^{-1}\left[x_{t+1}\ -\ c\omega_{t+1}\right]\ +\ C(I-\Lambda)m\,\omega_{t+1}\ +c\,\omega_t. \end{aligned}$$

But this relation is in fact equivalent to (B.4), given that our definitions above imply that

$$\begin{aligned} C\left[(I-\Lambda)M\ +\ \Lambda\right]C^{-1}\ &=\ B, \\ C\,(I-\Lambda)m\ &=\ Bc\ +\ b\cdot[-\beta\sigma^{-1}\ \ 0], \\ c\ &=\ b\cdot[\sigma^{-1}\ \ -1]. \end{aligned} \tag{B.6}$$

B.9 Derivation equation (3.6)-(3.9)

Since from $t \geq T$ $\bar{\imath}_t = \bar{\imath}_{LR}$ and by asumption $g_t = 0$ for all t, equation (2.13) can be written as:

$$\dot{e}_{LR} = [M-I]e_{LR} + m_2\bar{\imath}_{LR}$$

Since, by equation (2.11) $e_i^* = a_i$, where the equation for a is given by (2.12) replacing $m_2\bar{\imath}_{LR}$ instead of $m\omega$ since the first term in ω is 0. If $M-I$ is invertible, the unique rest point of this system is given by (3.6), which is calculated using the previous equation setting $\dot{e}_{LR} = 0$.

Given that the beliefs are started by $e_{LR}(0) = 0$, which are the ones consistent with the steady state in which the inflation target π^* is achieved at all times, and assuming that $M-I$ is not singular, we can write the solution for general n as[47]

$$e_{LR}(n) = [I - \exp[n(M-I)]]\bar{e}_{LR} \tag{B.7}$$

for all $n \geq 0$. As shown in section C, when the Taylor Principle (3.1) is satisfied, both

[47] See Hirsch and Smale (1974, p. 90).

eigenvalues of $M - I$ have negative real parts, and

$$\lim_{n\to\infty} \exp[n(M-I)] = 0 \tag{B.8}$$

It then follows (3.7).

For the periods before T, as stated in the text, we can calculate backwards the solution for any $t < T$, which depends on $\tau = T - t$. To do that, use (B.3) to get

$$\begin{aligned} e_t^* &= (I-\Lambda)\sum_{j=1}^{\infty}\Lambda^{j-1}\left[Me_{t+j} + m\omega_{t+j}\right] \\ &= (I-\Lambda)\sum_{j=t+1}^{T-1}\Lambda^{j-t-1}\left[Me_{t+j} + m_2\bar{\imath}_{SR}\right] + \sum_{j=T}^{\infty}\Lambda^{j-T-1}\left[Me_{LR} + m_2\bar{\imath}_{LR}\right] \end{aligned}$$

We can also write this equivalently for e_τ for $\tau \geq 1$, where $\tau = T - t$ as:

$$e_\tau^* = (I-\Lambda)\sum_{j=1}^{\tau-1}\Lambda^{j-1}\left[Me_{\tau-j} + m_2\bar{\imath}_{SR}\right] + \Lambda^{\tau-1}\left[Me_{LR} + m_2\bar{\imath}_{LR}\right]$$

Using this, now we can write the differential equation (2.13) as

$$e_\tau^*(n) = -e_\tau(n) + (I-\Lambda)\sum_{j=1}^{\tau-1}\Lambda^{j-1}\left[Me_{\tau-j} + m_2\bar{\imath}_{SR}\right] + \Lambda^{\tau-1}\left[Me_{LR} + m_2\bar{\imath}_{LR}\right] \tag{B.9}$$

and integrate forward from $e_\tau = 0$ for all $\tau \geq 1$ using the above solution for $e_{LR}(n)$. This is done by first solving for $\tau = 1$ uniquely given $e_{LR}(n)$, then for $\tau = 2$ and so on.

Equations (3.8) and (3.9) are the same as (B.7) and (3.6) just replacing LR by SR, since these equations are the behavior when the short run policy becomes permanent.

C Properties of Matrices

C.1 Properties of the Matrix M

A number of results turn on the eigenvalues of the matrix

$$M - I = \frac{1}{\Delta}\begin{bmatrix} -\frac{\sigma\phi_y + \sigma\kappa\phi_\pi - \sigma\kappa}{1-\beta} & \frac{(1-\alpha)\sigma\beta(1+\sigma\phi_y-\phi_\pi)}{1-\beta} \\ \frac{\kappa}{(1-\alpha)(1-\alpha\beta)} & \frac{\beta(1+\sigma\phi_y)-\Delta}{1-\alpha\beta} \end{bmatrix}.$$

We first note that the determinant of the matrix is given by

$$\text{Det}(M-I) \;=\; \frac{\sigma\kappa}{\Delta(1-\beta)(1-\alpha\beta)}\left(\phi_\pi + \frac{(1-\beta)}{\kappa}\phi_y - 1\right).$$

Under our sign assumptions, the factor pre-multiplying the factor in parentheses is necessarily positive. Hence the determinant is non-zero (and the matrix is non-singular) if

$$\phi_\pi \;+\; \frac{(1-\beta)}{\kappa}\phi_y \;-\; 1 \;\neq\; 0. \tag{C.1}$$

(In this case the steady-state vector of expectations (3.6) is well-defined, as asserted in the text.)

For any 2×2 real matrix A, both eigenvalues have negative real part if and only if $\text{Det}[A] > 0$ and $\text{Tr}[A] < 0$.[48] From the result above, the first of these conditions is satisfied if the left-hand side of (C.1) is positive, which is to say, if the Taylor Principle (3.1) is satisfied. The trace of $M - I$ is given by

$$\text{Tr}(M-I) \;=\; -\frac{1}{\Delta}\left(\frac{\sigma(\phi_y + \kappa\phi_\pi - \kappa)}{1-\beta} + \frac{\sigma\kappa\phi_\pi + (1-\beta)(1+\sigma\phi_y)}{1-\alpha\beta}\right).$$

The second term inside the parentheses is necessarily positive under our sign assumptions, and the first term is positive as well if the Taylor Principle is satisfied, since

$$\phi_y + \kappa\phi_\pi - \kappa \;=\; \kappa\left(\phi_\pi + \frac{\phi_y}{\kappa} - 1\right) \;>\; \kappa\left(\phi_\pi + \frac{\phi_y(1-\beta)}{\kappa} - 1\right) \;>\; 0. \tag{C.2}$$

Hence the Taylor Principle is a sufficient condition for $\text{Tr}[M - I] < 0$. It follows that (given our other sign assumptions) the Taylor Principle is both necessary and sufficient for both eigenvalues of $M - I$ to have negative real part.

If instead the left-hand side of (C.1) is negative, $\text{Det}[M - I] < 0$, and as a consequence the matrix must have two real eigenvalues of opposite sign.[49] Thus one eigenvalue is positive in this case, as asserted in the text. Note that this is the case that obtains if $\phi_\pi = \phi_y = 0$. Note also that in the case that $\phi_y = 0$, the condition becomes $\phi_\pi > 1$, which is the assumption in the Simple Illustration.

[48] See, for example, Hirsch and Smale (1974, p. 96).

[49] Again see Hirsch and Smale (1974, p. 96).

C.2 A Further Implication of the Taylor Principle

We are also interested in the eigenvalues of the related matrix $A(\lambda)M - I$, where for an arbitrary real number $-1 \leq \lambda \leq 1$, we define

$$A(\lambda) \equiv \begin{pmatrix} \frac{\lambda(1-\delta_1)}{1-\lambda\delta_1} & 0 \\ 0 & \frac{\lambda(1-\delta_2)}{1-\lambda\delta_2} \end{pmatrix}.$$

(Note that in the limiting case $\lambda = 1$, this reduces to the matrix $M - I$, just discussed.) In the case that the Taylor principle (3.1) is satisfied, we can show that for any $-1 \leq \lambda \leq 1$, both eigenvalues of $A(\lambda)M - I$ have negative real part. This follows again from a consideration of the determinant and trace of the matrix (generalizing the above discussion).

Since

$$A(\lambda)M - I = \frac{1}{\Delta}\begin{bmatrix} -\frac{\Delta-\lambda(1+\sigma\kappa)}{1-\beta\lambda} & -\frac{\sigma(1-\alpha)\beta(\phi_\pi-1-\sigma\phi_y)\lambda}{1-\beta\lambda} \\ \frac{\kappa\lambda}{(1-\alpha)(1-\alpha\beta\lambda)} & -\frac{\Delta-\beta\lambda(1+\sigma\phi_y)}{1-\alpha\beta\lambda} \end{bmatrix},$$

we have

$$\text{Det}(A(\lambda)M - I) = \frac{\Delta - \lambda(\beta(1+\sigma\phi_y) + 1 + \sigma\kappa) + \beta\lambda^2}{\Delta(1-\beta\lambda)(1-\alpha\beta\lambda)}.$$

Note that under our sign assumptions, the denominator is necessarily positive. The numerator defines a function $g(\lambda)$, a convex function (a parabola) with the properties

$$g'(1) = (\beta - 1) - \beta\sigma\phi_y - \kappa\sigma < 0$$

and

$$g(1) = \kappa\sigma\left(\phi_\pi + \frac{(1-\beta)}{\kappa}\phi_y - 1\right),$$

so that $g(1) > 0$ if and only if the Taylor Principle is satisfied. Hence the function $g(\lambda) > 0$ for all $\lambda \leq 1$, with the consequence that $\text{Det}[A(\lambda)M - I] > 0$ for all $|\lambda| \leq 1$, if and only if the Taylor Principle is satisfied.

The trace of the matrix is given by

$$\text{Tr}(A(\lambda)M - I) = -\frac{1}{\Delta}\left(\frac{\Delta - \lambda(1+\sigma\kappa)}{1-\beta\lambda} + \frac{\Delta - \beta\lambda(1+\sigma\phi_y)}{1-\alpha\beta\lambda}\right).$$

The denominators of both terms inside the parentheses are positive for all $|\lambda| \leq 1$, and we necessarily have $\Delta > 0$ under our sign assumptions as well. The numerator of

the first term inside the parentheses is also positive, since

$$\Delta - \lambda(1+\sigma\kappa) \;=\; \sigma\left[\kappa\phi_\pi + \phi_y - \kappa\right] \;+\; (1-\lambda)(1+\sigma\kappa) \;\geq \sigma\left[\kappa\phi_\pi + \phi_y - \kappa\right] \;>\; 0$$

if the Taylor Principle is satisfied, again using (C.2). And the numerator of the second term inside the parentheses is positive as well, since

$$\Delta - \beta\lambda(1+\sigma\phi_y) \;=\; (1-\beta\lambda)(1+\sigma\phi_y) \;+\; \kappa\sigma\phi_\pi \;>\; 0$$

under our sign assumptions. Thus the Taylor Principle is also a sufficient condition for $\mathrm{Tr}[A(\lambda)M - I] < 0$ for all $|\lambda| \leq 1$.

It then follows that the Taylor Principle is necessary and sufficient for both eigenvalues of the matrix $A(\lambda)M - I$ to have negative real part, in the case of any $|\lambda| < 1$. We use this result in the proof of Proposition 1.

C.3 Properties of the Matrix B

Necessary and sufficient conditions for both eigenvalues of a 2×2 matrix B to have modulus less than 1 are that (i) $\mathrm{Det}B < 1$; (ii) $\mathrm{Det}B + \mathrm{Tr}B > -1$; and (iii) $\mathrm{Det}B - \mathrm{Tr}B > -1$. In the case of the matrix B defined above, we observe that

$$\Delta\,\mathrm{Det}B \;=\; \beta, \tag{C.3}$$

$$\Delta\,\mathrm{Tr}B \;=\; 1 \;+\; \kappa\sigma \;+\; \beta(1+\sigma\phi_y).$$

From these facts we observe that our general sign assumptions imply that

$$\Delta\,\mathrm{Det}B \;<\; \Delta,$$

$$\Delta\,(\mathrm{Det}B \;+\; \mathrm{Tr}B \;+1) \;>\; 0.$$

Thus (since Δ is positive) conditions (i) and (ii) from the previous paragraph necessarily hold. We also find that

$$\Delta\,(\mathrm{Det}B \;-\; \mathrm{Tr}B \;+1) \;=\; \kappa\sigma\left[\phi_\pi \;+\; \left(\frac{1-\beta}{\kappa}\right)\phi_y \;-\; 1\right],$$

from which it follows that condition (iii) is also satisfied if and only if the quantity in the square brackets is positive. Thus we conclude that both eigenvalues of B have modulus less than 1 if and only if the Taylor Principle (3.1) is satisfied.

In the case that the Taylor Principle is violated (as in the case of a fixed interest rate, in which case $\phi_\pi = \phi_y = 0$), since $\mathrm{Det}B = \mu_1\mu_2$ and $\mathrm{Tr}B = \mu_1 + \mu_2$, where (μ_1, μ_2) are the two eigenvalues of B, the fact that condition (iii) fails to hold implies

that

$$(\mu_1 - 1)(\mu_2 - 1) \; < \; 0. \tag{C.4}$$

This condition is inconsistent with the eigenvalues being a pair of complex conjugates, so in this case there must be two real eigenvalues. Condition (C.4) further implies that one must be greater than 1, while the other is less than 1. Condition (C.3) implies that $\mathrm{Det}B \; > \; 0$, which requires that the two real eigenvalues both be non-zero and of the same sign; hence both must be positive. Thus when the Taylor Principle is violated (i.e., the quantity in (C.1) is negative), there are two real eigenvalues satisfying

$$0 \; < \; \mu_1 \; < \; 1 \; < \; \mu_2,$$

as asserted in section 2.2.

We further note that in this case, e_2', the (real) left eigenvector associated with eigenvalue μ_2, must be such that $e_2'b \neq 0$ (a result that is relied upon in section 4.2). The vector $v_2' \neq 0$ must satisfy

$$e_2' \left[B \; - \; \mu_2 I\right] \; = \; 0$$

to be a left eigenvector. The first column of this relation implies that $(1 - \mu_2)e_{2,1} \; + \; \kappa e_{2,2} \; = \; 0$, where we use the notation $e_{2,j}$ for the jth element of eigenvector e_2'. Since $\kappa > 0$ and $\mu_2 > 1$, this requires that $e_{2,1}$ and $e_{2,2}$ must both be non-zero and have the same sign. But since both elements of b have the same sign, this implies that $e_2'b \neq 0$.

Finally, we note that whenever (C.1) holds, regardless of the sign, the eigenvalues must satisfy

$$(\mu_1 - 1)(\mu_2 - 1) \; \neq \; 0,$$

so that B has no eigenvalue equal exactly to 1. This means that the matrix $B - I$ must be non-singular, which is the condition needed for existence of unique steady-state levels of output and inflation consistent with a PFE. In the case of constant fundamentals $\omega_t = \bar{\omega}$ for all t, the unique steady-state solution to (B.4) is then given by $x_t = \bar{x}$ for all t, where

$$\bar{x} \; \equiv \; (I - B)^{-1} \, b \, [(1 - \beta)\sigma^{-1}\bar{g} \; - \; \bar{\imath}]. \tag{C.5}$$

Note that condition (C.1) is also the condition under which $M - I$ is non-singular, as shown above. Moreover, since $I - \Lambda$ is non-singular, $M - I$ is non-singular if and only if $(I - \Lambda)(M - I) \; = \; [(I - \Lambda)M \; + \; \Lambda] \; - \; I$ is non-singular. This is the condition under which equation (B.5) has a unique steady-state solution, in which $e_t = \bar{e}$ for all t, with

$$\bar{e} \; \equiv \; (I - M)^{-1} \, m \, \bar{\omega}.$$

This solution for steady-state PFE expectations is consistent with (C.5) because of

the identities linking the M and B matrices noted above.

D Proofs of Propositions

D.1 Proof of Proposition 1

Under the hypotheses of the proposition, there must exist a date $\bar{T}$ such that the fundamental disturbances $\{\omega_t\}$ can be written in the form

$$\omega_t \;=\; \omega_\infty + \sum_{k=1}^{K} a_{\omega,k} \lambda_k^{t-\bar{T}}$$

for all $t \geq \bar{T}$, and the initial conjecture can also be written in the form

$$e_t(0) \;=\; e_\infty(0) + \sum_{k=1}^{K} a_{e,k}(0) \lambda_k^{t-\bar{T}}$$

for all $t \geq \bar{T}$, where $|\lambda_k| < 1$ for all $k = 1, \ldots, K$. (There is no loss of generality in using the same date $\bar{T}$ and the same finite set of convergence rates $\{\lambda_k\}$ in both expressions.) With a driving process and initial condition of this special form, the solution to the system of differential equations (2.13) will be of the form

$$e_t(n) \;=\; e_\infty(n) + \sum_{k=1}^{K} a_{e,k}(n) \lambda_k^{t-\bar{T}}$$

for all $t \geq \bar{T}$, for each $n \geq 0$. We then need simply determine the evolution as n increases of the finite set of values $e_t(n)$ for $0 \leq t \leq T-1$, together with the finite set of coefficients $e_\infty(n)$ and $a_{e,k}(n)$. This is a set of $2(\bar{T} + K + 1)$ functions of n, which we write as the vector-valued function $\mathbf{e}(n)$ in the text.

In the case of any belief sequences and disturbances of the form assumed in the above paragraph, it follows from (B.3) that the implied correct beliefs will be of the form

$$e_t^*(n) \;=\; e_\infty^*(n) + \sum_{k=1}^{K} a_{e,k}^*(n) \lambda_k^{t-\bar{T}}$$

for all $t \geq \bar{T}$, where

$$e_\infty^*(n) \;=\; M\, e_\infty(n) \;+\; m\, \omega_\infty,$$

and

$$a_{e,k}^*(n) \;=\; A(\lambda_k)\, [M\, a_{e,k}(n) \;+\; m\, a_{\omega,k}]$$

for each $k = 1, \ldots, K$. We further observe that for any $t < \bar{T}$,

$$e_t^*(n) \;=\; (I - \Lambda) \sum_{j=1}^{\bar{T}-t-1} \Lambda^j [M e_{t+j}(n) + m\omega_{t+j}] \;+\; \Lambda^{\bar{T}-t-1} [M e_\infty(n) \,+\, m\omega_\infty]$$
$$+ \;\sum_{k=1}^{K} \lambda_k^{-1} \Lambda^{\bar{T}-t-1} A(\lambda_k) [M a_{e,k}(n) \,+\, m a_{\omega,k}].$$

Thus the sequence $\{e_t^*(n)\}$ can also be summarized by a set of $2(\bar{T} + K + 1)$ functions of n, and each of these is a linear function of the elements of the vectors $\mathbf{e}(n)$ and $\boldsymbol{\omega}$.

It then follows that the dynamics (2.13) can be written in the more compact form

$$\dot{\mathbf{e}}(n) \;=\; V\,\mathbf{e}(n) \;+ W\,\boldsymbol{\omega}, \tag{D.1}$$

where the elements of the matrices V and W are given by the coefficients of the equations in the previous paragraph. Suppose that we order the elements of $\mathbf{e}(n)$ as follows: the first two elements are the elements of e_0, the next two elements are the elements of e_1, and so on, through the elements of $e_{\bar{T}-1}$; the next two elements are the elements of $a_{e,1}$, the two elements after that are the elements of $a_{e,2}$, and so on, through the elements of $a_{e,K}$; and the final two elements are the elements of e_∞. Then we observe that the matrix V is of the form

$$V \;=\; \begin{bmatrix} V_{11} & V_{12} \\ 0 & V_{22} \end{bmatrix}, \tag{D.2}$$

where the first $2\bar{T}$ rows are partitioned from the last $2(K+1)$ rows, and the columns are similarly partitioned.

Moreover, the block V_{11} of the matrix is of the block upper-triangular form

$$V_{11} \;=\; \begin{bmatrix} -I & v_{12} & \cdots & v_{1,\bar{T}-1} & v_{1,\bar{T}} \\ 0 & -I & \cdots & v_{2,\bar{T}-1} & v_{2,\bar{T}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -I & v_{\bar{T}-1,\bar{T}} \\ 0 & 0 & \cdots & 0 & -I \end{bmatrix}, \tag{D.3}$$

where now each block of the matrix is 2×2. Furthermore, when V_{22} is similarly

partitioned into 2×2 blocks, it takes the block-diagonal form

$$V_{22} \;=\; \left[\begin{array}{cccc} A(\lambda_1)M - I & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & A(\lambda_K)M - I & 0 \\ 0 & \cdots & 0 & M - I \end{array}\right]. \tag{D.4}$$

These results allow us to determine the eigenvalues of V. The block-triangular form (D.2) implies that the eigenvalues of V consist of the $2\bar{T}$ eigenvalues of V_{11} and the $2(K+1)$ eigenalues of V_{22} (the two diagonal blocks). Similarly, the block-triangular form (D.3) implies that the eigenvalues of V_{11} consist of the eigenvalues of the diagonal blocks (each of which is $-I$), which means that the eigenvalue -1 is repeated $2\bar{T}$ times. Finally, the block-diagonal form (D.4) implies that the eigenvalues of V_{22} consist of the eigenvalues of the diagonal blocks: the two eigenvalues of $A(\lambda_k)M - I$, for each $k = 1, \ldots, K$, and the two eigenvalues of $M - I$.

Using the results in section C.1, it follows from the hypothesis that the reaction function coefficients satisfy (3.1) and the hypothesis that $|\lambda_k| < 1$ for each k that all of the eigenvalues of $M - I$ and of each of the matrices $A(\lambda_k)M - I$ have negative real part. Since all of the other eigenvalues of V are equal to -1, all $2(\bar{T} + K + 1)$ eigenvalues of V have negative real part. This implies that V is non-singular, so that there is a unique rest point for the dynamics (D.1), defined by:

$$\mathbf{e}^{PF} \;\equiv\; -V^{-1}W\,\boldsymbol{\omega}.$$

It also implies that the dynamics (D.1) converge asymptotically to that rest point as n goes to infinity, for any initial condition $\mathbf{e}(0)$ (Hirsch and Smale, 1974, pp. 90-95).[50]

The rest point to which $\mathbf{e}(n)$ converges is easily seen to correspond to the unique PFE that belongs to the same linear space L^2. Beliefs in L^2 constitute a PFE if and only if $\mathbf{e}^* = \mathbf{e}$. From our characterization above of $\mathbf{e}^*$, this is equivalent to the requirement that $V\,\mathbf{e} + W = 0$, which holds if and only if $\mathbf{e} = \mathbf{e}^{PF}$, the unique rest point of the system (D.1).

Finally, the paths of output and inflation in any reflective equilibrium are given by (2.10), given the solution for $\{e_t(n)\}$. Using (2.9), one obtains a similar linear equation for the nominal interest rate each period. It then follows that for any t, the reflective equilibrium values for y_t, π_t, and i_t converge to the FS-PFE values as n is made large.

[50] Of course, it is important to recognize that this result only establishes convergence for initial conjectures that belong to the linear space L^2. The result also only establishes convergence under the assumption that the linear dynamics (D.1) apply at all times; this depends on assuming that the reaction function (2.9) can be implemented at all times, which requires that the ZLB never bind. Thus we only establish convergence for all those initial conjectures such that the dynamics implied by (2.13) never cause the ZLB to bind. There is however a large set of initial conditions for which this is true, given that the unconstrained dynamics are asymptotically convergent.

Furthermore, the complete sequences of values for these three variables for any value of n depend on only the finite number of elements of the vector $\mathbf{e}(n)$, in such a way that for any $\epsilon > 0$, there exists an $\tilde{\epsilon} > 0$ such that it is guaranteed that each of the variables y_t, π_t, and i_t are within distance ϵ of their FS-PFE values for all t as long as $|\mathbf{e}(n) - \mathbf{e}^{PF}| < \tilde{\epsilon}$. The convergence of $\mathbf{e}(n)$ to $\mathbf{e}^{PF}$ then implies the existence of a finite $n(\epsilon)$ for which the latter condition is satisfied, regardless of how small $\tilde{\epsilon}$ needs to be. This proves the proposition.

D.2 Proof of Proposition 2

It has already been shown in the text that under the assumptions of the proposition, we have $e_t(n) = e_{LR}(n)$ for all $t \geq T$, where $e_{LR}(n)$ is given by (B.7). It has also been shown that for any $\tau \geq 1$, the solution for $e_\tau(n)$, where $\tau \equiv T - t$ is the number of periods remaining until the regime change, is independent of T. The functions $\{e_\tau(n)\}$ further satisfy the system of differential equations

$$\begin{aligned} \dot{e}_\tau(n) \;=\; -e_\tau(n) \;&+\; (I - \Lambda) \sum_{j=1}^{\tau-1} \Lambda^{j-1} \left[M e_{\tau-j}(n) \,+\, m_2 \bar{\imath}_{SR} \right] \\ &+\; \Lambda^{\tau-1} \left[M e_{LR}(n) \,+\, m_2 \bar{\imath}_{LR} \right] \end{aligned} \tag{D.5}$$

derived in the text, together with the initial conditions $e_\tau(0) = 0$ for all $\tau \geq 1$. (Equation (D.5) repeats equation (B.9) from the text.)

We wish to calculate the behavior of the solution to this system as $\tau \to \infty$ for an arbitrary value of n. It is convenient to use the method of z-transforms (Jury, 1964). For any n, let the z-transform of the sequence $\{e_\tau(n)\}$ for $\tau \geq 1$ be defined as the function

$$X_n(z) \;\equiv\; \sum_{\tau=1}^{\infty} e_\tau(n)\, z^{1-\tau}. \tag{D.6}$$

Here $X_n(z)$ is a vector-valued function; each element is a function of the complex number z, defined for complex numbers $|z| > 1/\rho$, where ρ is the minimum of the radii of the convergence of the two series.

Differentiating (D.6) with respect to n, and substituting (D.5) for $\dot{e}_\tau(n)$ in the

resulting equation, we obtain an evolution equation for the z-transform:

$$\begin{aligned}\dot{X}_n(z) &= -\sum_{\tau=1}^{\infty} e_\tau(n) z^{1-\tau} + (I-\Lambda)\sum_{j=0}^{\infty}\Lambda^j z^{-j}\left[M\sum_{\tau=1}^{\infty} e_\tau(n) z^{-\tau} + m_2\bar{\imath}_{SR}\sum_{\tau=1}^{\infty} z^{-\tau}\right] \\ &\qquad + \sum_{j=0}^{\infty}\Lambda^j z^{-j}\left[M\, e_{LR}(n) + m_2\bar{\imath}_{LR}\right] \\ &= -X_n(z) + (I-\Lambda)(I-\Lambda z^{-1})^{-1}\left[z^{-1}M\, X_n(z) + (z-1)^{-1} m_2\bar{\imath}_{SR}\right] \\ &\qquad + (I-\Lambda z^{-1})^{-1}\left[M e_{LR}(n) + m_2\bar{\imath}_{LR}\right], \end{aligned} \tag{D.7}$$

which holds for any $n > 0$ and any z in the region of convergence. We note that the right-hand side of (D.7) is well-defined for all $|z| > 1$.

The z-transform of the initial condition is simply $X_0(z) = 0$ for all z. Thus we wish to find functions $\{X_n(z)\}$ for all $n \geq 0$, each defined on the region $|z| > 1$, that satisfy (D.7) for all n and all $|z| > 1$, together with the initial condition $X_0(z) = 0$ for all z. If we can find such a solution, then for any n we can find the implied sequence $\{e_t(n)\}$ by inverse z-transformation of the function $X_n(z)$.

We note that the dynamics of $X_n(z)$ implied by (D.7) is independent for each value of z. (This is the advantage of z-transformation of the original system of equations (D.5).) Thus for each value of z such that $|z| > 1$, we have an independent first-order ordinary differential equation to solve for $X_n(z)$, with the single initial condition $X_0(z) = 0$. This equation has a closed-form solution for each z, given by

$$\begin{aligned} X_n(z) &= (1-z^{-1})^{-1}\left[I - \exp(n(M-I))\right](I-M)^{-1}\cdot m_2\bar{\imath}_{LR} \\ &\quad + (z-1)^{-1}\left[I - \exp(-n\Phi(z))\right]\Phi(z)^{-1}(I-\Lambda)(I-\Lambda z^{-1})^{-1} \\ &\qquad \cdot m_2(\bar{\imath}_{SR} - \bar{\imath}_{LR}) \end{aligned} \tag{D.8}$$

for all $n \geq 0$, where

$$\Phi(z) \equiv I - (I-\Lambda)(I-\Lambda z^{-1})^{-1} z^{-1} M.$$

Note also that the expression on the right-hand side of (D.8) is an analytic function of z everywhere in the complex plane outside the unit circle, and can be expressed as a sum of powers of z^{-1} that converges everywhere in that region. Such a series expansion of $X_n(z)$ for any n allows us to recover the series of coefficients $\{e_\tau(n)\}$ associated with the reflective equilibrium with degree of reflection n.

For any value of $n \geq 0$, we are interested in computing

$$e_{SR}(n) \equiv \lim_{T\to\infty} e_t(n) = \lim_{\tau\to\infty} e_\tau(n).$$

The final value theorem for z-transforms[51] implies that

$$\lim_{\tau\to\infty} e_\tau(n) \;=\; \lim_{z\to 1}(z-1)X_n(z)$$

if the limit on the right-hand side exists. In the case of the solution (D.8), we observe that the limit is well-defined, and equal to

$$\lim_{z\to 1}(z-1)X_n(z) \;=\; [I \;-\; \exp(n(M-I))]\;(I-M)^{-1}\,m_2\bar{\imath}_{SR}.$$

Hence for any t and any n, $e_t(n)$ converges to a well-defined (finite) limit as T is made large, and the limit is the one given in the statement of the proposition.

D.3 Proof of Proposition 3

The result that

$$\lim_{T\to\infty} e_t(n) \;=\; e_{SR}(n)$$

for all t and n follows from Proposition 2. If in addition, the Taylor Principle (3.1) is satisfied, then as shown in section C.1 above, both eigenvalues of $M-I$ have negative real part. From this (B.8) follows; substituting of this into (3.8) yields

$$\lim_{n\to\infty} e_{SR}(n) \;=\; \bar{e}^{PF}_{SR},$$

where $\bar{e}^{PF}_{SR}$ is defined in (3.9). This establishes the first double limit in the statement of the proposition.

The result that

$$\lim_{n\to\infty} e_t(n) \;=\; e^{PF}_t$$

for all t follows from Proposition 1. Establishing the second double limit thus requires us to consider how e^{PF}_t changes as T is made large.

As discussed in section B.8 above, the FS-PFE dynamics $\{e^{PF}_t\}$ satisfy equation (B.5) for all t. Under the kind of regime assumed in this proposition (with ω_t equal to a constant vector $\bar{\omega}$ for all $t \geq T$), the FS-PFE (obtained by "solving forward" the difference equation) involves a constant vector of expectations, $e^{PF}_t = \bar{e}^{PF}_{LR}$ for all $t \geq T-1$, where

$$\bar{e}^{PF}_{LR} \equiv [I-M]^{-1}\,m_2\bar{\imath}_{LR}$$

is the same as the vector defined in (3.6).

For periods $t < T-1$, one must instead solve the difference equation backward from the terminal condition $e^{PF}_{T-1} = \bar{e}^{PF}_{LR}$. We thus obtain a difference equation of the

[51] See, for example, Jury (1964, p. 6).

form

$$e_\tau \;=\; [(I-\Lambda)M \;+\; \Lambda]\, e_{\tau-1} \;+\; (I-\Lambda)\, m_2 \bar{\imath}_{SR} \tag{D.9}$$

for all $\tau \geq 2$, with initial condition $e_1 = \bar{e}^{PF}_{LR}$. The asymptotic behavior of these dynamics as τ is made large depends on the eigenvalues of the matrix

$$(I-\Lambda)M \;+\; \Lambda \;=\; C^{-1}BC, \tag{D.10}$$

which must be the same as the eigenvalues of B. (Note that (D.10) follows from (B.6).)

Under the hypothesis that the response coefficients satisfy the Taylor Principle (3.1), both eigenvalues of B are inside the unit circle. It then follows that the dynamics (D.9) converge as $\tau \to \infty$ to the steady-state vector of expectations $\bar{e}^{PF}_{SR}$ defined in (3.9). We thus conclude that

$$\lim_{T\to\infty} e^{PF}_t \;=\; \bar{e}^{PF}_{SR}$$

for any t. This establishes the second double limit.

D.4 Proof of Proposition 4

The proof of this proposition follows exactly the same lines as the proof of Proposition 1. While the definition of the matrices of coefficients V and W must be modified, it continues to be possible to write the belief revision dynamics in the compact form (D.1), for an appropriate definition of these matrices. (This depends on the fact that we have chosen $\bar{T} \geq T$, so that the coefficients of the monetary policy reaction function do not change over time during periods $t \geq \bar{T}$. Variation over time in the reaction function coefficients does not prevent us from writing the dynamics in the compact form, as long as it occurs only prior to date $\bar{T}$; and our method of analysis requires only that $\bar{T}$ be finite.)

Moreover, it continues to be the case that V will have the block-triangular form indicated in equations (D.2)–(D.4). In equation (D.4), the matrix M is defined using the coefficients (ϕ_π, ϕ_y) that apply after date T, and thus that satisfy the Taylor Principle (3.1), according to the hypotheses of the proposition. The eigenvalues of V again consist of -1 (repeated $2\bar{T}$ times); the eigenvalues of $A(\lambda_k)M$, for $k = 1, \ldots, K$, and the eigenvalues of M. Because M is defined using coefficients that satisfy the Taylor Principle, we again find that all of the eigenvalues of M and of $A(\lambda_k)M$ have negative real part. Hence all of the eigenvalues of V have negative real part. This again implies that the dynamics (D.1) are asymptotically stable, and the fixed point to which they converge again corresponds to the FS-PFE expectations. This establishes the proposition.

Note that this result depends on the hypothesis that from date T onward, monetary policy is determined by a reaction function with coefficients that satisfy the Taylor Principle. If we assumed instead (as in the case emphasized in Cochrane, 2017) that

after date T, policy again consists of a fixed interest rate, but one that is consistent with the long-run inflation target (i.e., $\bar{\imath}_{LR} = 0$), the belief-revision dynamics would *not* converge. (See the discussion in section 4.3 of the text of the case in which an interest-rate peg differs temporarily from the long-run interest-rate peg.)

If the interest rate is also fixed after date T (albeit at some level $\bar{\imath}_{LR} \neq \bar{\imath}_{SR}$), the belief-revision dynamics can again be written in the compact form (D.1), and the matrix V will again have the form (D.2)–(D.4). But in this case, the matrix M in (D.4) would be defined using the response coefficients $\phi_\pi = \phi_y = 0$, so that the Taylor Principle is violated. It then follows from our results above that M will have a positive real eigenvalue. (By continuity, one can show that $A(\lambda_k)M$ will also have a positive real eigenvalue for all values of λ_k near enough to 1.) Hence V will have at least one (and possibly several) eigenvalues with positive real part, and the belief-revision dynamics (D.1) will be explosive in the case of almost all initial conjectures (even restricting our attention to conjectures within the specified finite-dimensional family).

D.5 Proof of Proposition 5

The proof of this proposition follows similar lines as the proof of Proposition 2. In general, the characterization of reflective equilibrium is more complex when the monetary policy response coefficients are not time-invariant, as in the situation considered here. However, in the case hypothesized in the proposition, $g_t = 0$ and from period T onward, monetary policy is consistent with constant inflation at the rate π^*. Under these circumstances, and initial conjecture under which $e_t = 0$ for all $t \geq T$ implies correct beliefs $e_t^* = 0$ for all $t \geq T$ as well. Hence under the belief-revision dynamics, the conjectured beliefs are never revised, and $e_t(n) = 0$ for all degrees of reflection $n \geq 0$, and any $t \geq T$. This result would be *the same* if we were to assume a fixed interest rate for all $t \geq T$ (that is, if we were to assume response coefficients $\phi_\pi = \phi_y = 0$ after date T, just like we do for dates prior to T), but a fixed interest rate $\bar{\imath}_t = 0$ for all $t \geq T$ (that is, the fixed interest rate consistent with the steady state with inflation rate π^*).

Thus the reflective equilibrium is the same (in this very special case) as if we assumed a fixed interest rate in all periods (and thus the same response coefficients in all periods), but $\bar{\imath}_t = \bar{\imath}_{SR}$ for $t < T$ while $\bar{\imath}_t = 0$ for $t \geq T$.[52] And the latter is a case to which Proposition 2 applies. (Note that Proposition 2 requires no assumptions about the response coefficients except that they are constant over time, and that they

[52]Note that these two different specifications of monetary policy would *not* lead to the same reflective equilibrium expectations, under most assumptions about the real shocks or about the initial conjecture; see the discussion at the end of the proof of Proposition 4. Here we get the same result *only* because we assume $g_t = 0$ (exactly) for all $t \geq T$ and an initial conjecture under which $e_t(0) = 0$ (exactly) for all $t \geq T$.

satisfy (C.1). Hence the case in which $\phi_\pi = \phi_y = 0$ in all periods is consistent with the hypotheses of that proposition.)

Proposition 2 can then be used to show that the reflective equilibrium beliefs $\{e_t(n)\}$ for any degree of reflection n converge to a well-defined limiting value $e_{SR}(n)$, which is given by (3.8)–(3.9). This establishes the proposition.

D.6 Proof of Proposition 6

Let $\{e_t^1\}$ be the sequence of expectations in a reflective equilibrium when the date of the regime change is T, and $\{e_t^2\}$ be the expectations in the equilibrium corresponding to the same degree of reflection n when the date of the regime change is $T' > T$. Similarly, let $\{a_t^1\}$ and $\{a_t^2\}$ be the evolution of the vectors of summary variables that decisionmakers need to forecast in the two equilibria, and $\{e_t^{*1}\}$ and $\{e_t^{*2}\}$ the implied sequences of correct forecasts in the two equilibria. We similarly use the notation $M^{(i)}, m^{(i)}, C^{(i)}, c^{(i)}$ to refer to the matrices M, m, C, and c respectively, defined using the monetary policy response coefficients associated with regime i (for $i = 1, 2$).[53]

Let us first consider the predictions regarding reflective equilibrium in periods $t \geq T'$. Under both of the assumptions about policy, policy is expected to be the same at all dates $t \geq T'$. Since it is assumed that we start from the same initial conjecture $\{e_t(0)\}$ in both cases, and the model is purely forward-looking, it follows that the belief-revision dynamics will also be the same for all $t \geq T'$ in both cases. Hence we obtain the same sequences $\{e_t(n)\}$ in both cases, for all $t \geq T'$; and since the outcomes for output and inflation are then given by (2.10), these are the same for all $t \geq T'$ as well. Moreover, it is easily shown that under our assumptions, the common solution is one in which $e_t(n) = 0$ for all $t \geq T'$, and correspondingly $x_t(n) = 0$ for all $t \geq T'$.

Moreover, since outcomes for output and inflation are the same for all $t \geq T'$ in the two cases, it follows that the sequences of correct forecasts $\{e_t^*\}$ are the same in both cases for all $t \geq T' - 1$. (Note that the correct forecasts in period $T' - 1$ depend only on the equilibrium outcomes in period T' and later.) Hence the belief-revision dynamics for period $T' - 1$ will also be the same in both cases, and we obtain the same vector $e_{T'-1}(n)$ for all n; and again the common beliefs are $e_{T'-1}(n) = 0$.

Let us next consider reflective equilibrium in periods $T \leq t \leq T' - 1$. Suppose that for such t and some n, $e_t^2 \geq e_t^1 \geq 0$ (in both components). Then

$$a_t^2 - a_t^1 \;=\; M^{(2)}\,(e^2 - e_t^1) \;+\; [M^{(2)} - M^{(1)}]\,e_t^1 \,|\, +\; m_2^{(2)}\,\bar{\imath}_{SR}.$$

Moreover, we observe from the above definitions of M and m that $M^{(2)}$ is positive in all elements; $M^{(2)} - M^{(1)}$ is positive in all elements; and $m_2^{(2)}$ is negative in both elements. Under the hypotheses that $e_t^2 \geq e_t^1 \geq 0$ and $\bar{\imath}_{SR} < 0$, it follows that

[53] By "regime 1" we mean the Taylor rule (the regime in place in periods $T \leq t < T'$ under policy 1); by "regime 2" we mean the interest-rate peg at $\bar{\imath}_{SR}$.

$a_t^2 - a_t^1 >> 0$, where we use the symbol $>>$ to indicate that the first vector is greater in both elements.

Now suppose that for some n, $e_t^2 \geq e_t^1 \geq 0$ for all $T \leq t \leq T' - 1$. It follows from our conclusions above that these inequalities then must hold for all $t \geq T$. It also follows from the argument in the paragraph above that we must have $a_t^2 >> a_t^1$ for all $T \leq t \leq T' - 1$, along with $a_t^2 = a_t^1$ for all $t \geq T'$. This implies that $e_t^{*2}(n) >> e_t^{*1}(n)$ for all $T \leq t < T' - 1$, while $e_t^{*2}(n) = e_t^{*1}(n)$ for $t = T' - 1$.

The fact that $e_t^{*2}(n) = e_t^{*1}(n)$ for $t = T' - 1$ means that the belief-revision dynamics for period $T' - 1$ will again be the same in both cases, and we obtain the same vector $e_{T'-1}(n)$ for all n; and again the common beliefs are $e_{T'-1}(n) = 0$. For periods $T \leq t < T' - 1$, we continue to have $e_t^{*1}(n) = 0$ for all n, for the same reason as in the case of periods $t \geq T'$. But now the fact that we start from the common initial conjecture $e_t^2(0) = e_t^1(0) = 0$ implies that $e_t^{*2}(0) >> e_t^{*1}(0) = 0$ and hence $\dot{e}_t^2(0) >> \dot{e}_t^1(0) = 0$. This implies that for small enough $n > 0$, we will have $e_t^2(n) >> e_t^1(n) = 0$ for all $T \leq t < T' - 1$.

Moreover, for any n, as long as we continue to have $e_t^2(n) \geq e_t^1(n) = 0$ for all $t \geq T$, we will continue to have $e_t^{*2}(n) >> e_t^{*1}(n) = 0$ for all $T \leq t < T' - 1$. Since the belief-revision dynamics (2.13) imply that for any $n > 0$, $e_t(n)$ is an average of $e_t(0)$ and the vectors $e_t^*(\tilde{n})$ for values $0 \leq \tilde{n} < n$, as long as we have had $e_t^{*2}(\tilde{n}) >> 0$ for all $0 \leq \tilde{n} < n$, we will necessarily have $e_t^2(n) >> 0$. Thus we conclude by induction that $e_t^2(n) >> e_t^1(n) = 0$ for all $n > 0$, and any $T \leq t < T' - 1$.

The associated reflective equilibrium outcomes are given by (2.10) in each case. This implies that

$$x_t^2 - x_t^1 \;=\; C^{(2)}\,(e^2 - e_t^1) \;+\; [C^{(2)} - C^{(1)}]\,e_t^1 \,|\, +\; c_2^{(2)}\,\bar{\imath}_{SR}.$$

Note furthermore that all elements of $C^{(2)}$ are non-negative, with at least one positive element in each row; that all elements of $C^{(2)} - C^{(1)}$ are positive; and that all elements of $c_2^{(2)}$ are negative. Then the fact that $e_t^2(n) \geq e_t^1(n) = 0$ for all $T \leq t \leq T' - 1$ and $\bar{\imath}_{SR} < 0$ implies that $x_t^2 >> x_t^1$ for all $T \leq t \leq T' - 1$.

Finally, let us consider reflective equilibrium in periods $0 \leq t < T$. In these periods, the monetary policy is expected to be the same in both cases (the fixed interest rate). Suppose that for some such t and some n, $e_t^2 \geq e_t^1$. Then

$$a_t^2 - a_t^1 \;=\; M^{(2)}\,(e^2 - e_t^1) \;\geq\; 0,$$

because all elements of $M^{(2)}$ are positive. Since we have already concluded above that $a_t^2 >> a_t^1$ for all $T \leq t \leq T' - 1$, and that $a_t^2 = a_t^1$ for all $t \geq T'$, this implies that $e_t^{*2} >> e_t^{*1}$ for all $0 \leq t < T$.

We can then use an inductive argument, as above, to show that $e_t^2(n) >> e_t^1(n)$

for any $n > 0$, and any $0 \leq t < T$. It follows from this that

$$x_t^2 - x_t^1 \;=\; C^{(2)}\,(e^2 - e_t^1) \;>>\; 0$$

for any $n > 0$, and any $0 \leq t < T$, given that all elements of $C^{(2)}$ are non-negative, with at least one positive element in each row. This establishes the proposition.

E Comparison with a Discrete Model of Belief Revision

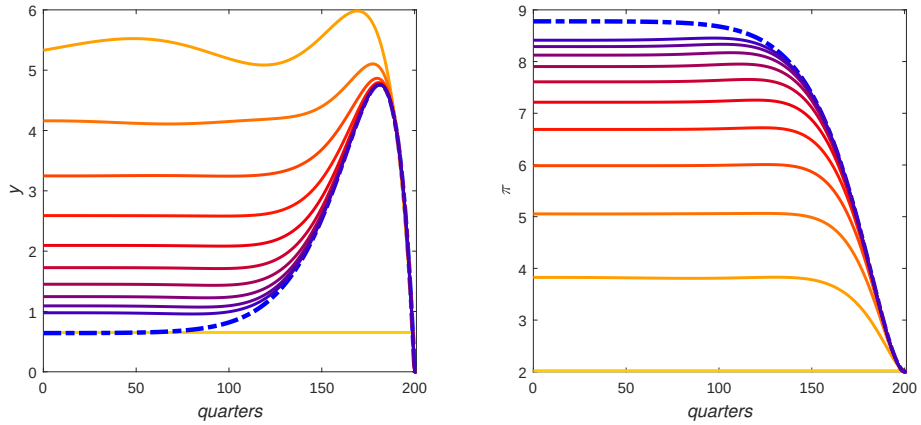
Here we note that the convergence result in Proposition 1 would not hold with the same generality were we instead to assume a discrete model of belief revision in which, instead of the continuous model of belief revision (2.13), we iterate the mapping

$$e_t(N+1) \;=\; e_t^*(N) \tag{E.1}$$

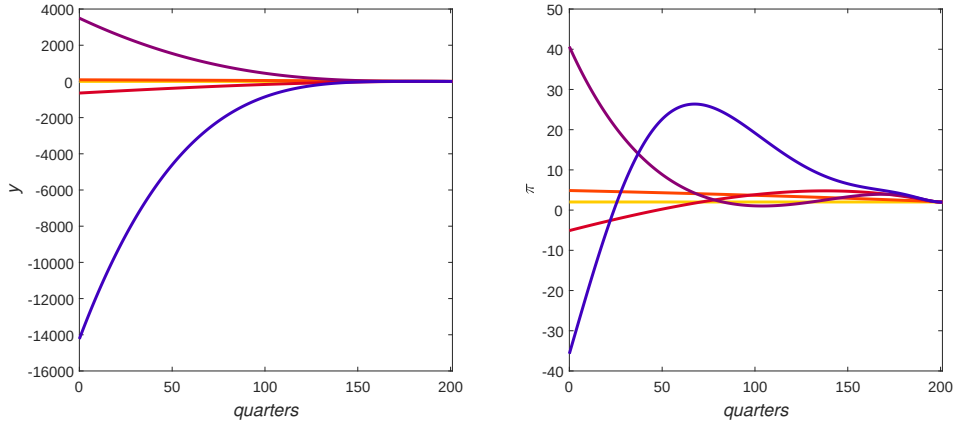
for $N = 0, 1, 2, \ldots$, where for each N, $\{e_t^*(N)\}$ is the sequence of correct beliefs implied by average expectations specified by the sequence $\{e_t(N)\}$. As with the continuous model, we might take as given some "naive" initial conjecture, and then consider how it evolves as a result of further iterations of the mapping. And as with the continuous model, *if* the process converges to a fixed point, such a fixed point must correspond to PFE beliefs.

Figure 5: Belief Revision Process using a Continuous vs. a Discrete process

(a) Continuous process



(b) Discrete process



Note: The graphs on top show the rsult for $n = 0$ through 20 (progressively darker lines) when the Taylor-rule intercept is reduced for 200 quarters. The graphs on the bottom show reflective equilibrium outcomes for $N = 0$ through 4 (progressively darker lines) when the Taylor-rule intercept is reduced for 200 quarters assuming a discrete process of iterative belief revision. See section F for details.

However, the conditions for convergence of the discrete process, while related to the conditions under which the continuous process converges, are more stringent. Convergence need not obtain under the conditions hypothesized in Proposition 1, as the following numerical example illustrates. In Figure 5 the intercept of the Taylor rule is expected to be lowered for 200 quarters, after which it is expected to return to the level consistent with the inflation target π^*. All model parameters are also the same as in Figure 2, and the initial conjecture is assumed to be $e_t(0) = 0$ for all t. In the panel on top, the continuous belief revision process is assumed and in the panel below,

the discretemodel of belief revision (E.1) is assumed.

The figure plots the implied TE dynamics of output and inflation for iterations $N = 0, 1, 2, 3$, and 4 for the discrete case. While the continuous case converges as is expected by Proposition 1, the belief-revision dynamics in the discrete case are explosive. The first revision of the initial conjecture (which takes account of the fact that it is predictable that if people maintain the initial beliefs, consistent with the unperturbed steady state, the temporary policy will lead to higher inflation and output) raises both output and inflation further. But anticipation of *these* effects (and the associated increase in the interest rate that they must provoke) should actually lead output and inflation to be *lower* in stage $N = 2$. Anticipation of the $N = 2$ outcomes (which imply an even deeper cut in the interest rate) then leads output and inflation to be *high* again in stage $N = 3$, and to an even greater extent than in stage $N = 1$. Anticipating of this then leads output and inflation to be *low* again in stage $N = 4$, to an even greater extent than in stage $N = 2$. The oscillations continue, growing larger and larger, as N is increased; but as the figure shows, the predicted expectations are already very extreme after only four iterations of the belief updating mapping.

It is not accidental that the unstable dynamics of belief revision in this case are oscillatory. In terms of the compact notation introduced in the proof of Proposition 1 (under the assumption of exponentially convergent fundamentals and average beliefs), the discrete model of belief revision (E.1) replaces the continuous dynamics (D.5) by the discrete process

$$\mathbf{e}(N+1) \;=\; (I+V)\,\mathbf{e}(N) \;+W\,\boldsymbol{\omega}$$

This process is unstable if not all eigenvalues of $I+V$ are of modulus less than 1. Since the eigenvalues of $I+V$ are equal to $1+\mu_i$, where μ_i is an eigenvalue of V, and we have shown above that all eigenvalues of V have negative real part, $I+V$ cannot have a real eigenvalue greater than 1. It can, however, have a real eigenvalue with *modulus* greater than 1, if V has a real eigenvalue that is less than -2. This is the case shown in Figure 5, in which a large negative eigenvalue results in explosive oscillations.

We feel, however, that the kind of unstable process of belief revision illustrated by Figure5 is unrealistic, as it is requires that at each stage in the reasoning, one must conjecture that *everyone* else should reason in one precise way, even though that assumed reasoning changes dramatically from each stage in the process of reflection to the next. The continuous process of belief revision proposed in the text avoids making such an implausible assumption.

F Algorithm to Construct the Figures

The figures were constructed using the parameters listed in Table 1.

Table 1: Parameters used in Numerical Exercises

Parameter	Definition	Value	Source
α	Prob. not choosing price	0.784	Denes et al. (2013)
β	Discount factor	0.997	
σ	Int. elast. substitution	1/1.22	
ξ	Elast. firm's optimal price wrt AD	0.125	
ϕ_y	Coef. output in Taylor rule	0.5/4	Taylor (1993)
ϕ_π	Coef. inflation in Taylor rule	1.5	
π^*	Inflation Target	$\log(1.02)^{1/4}$	

The initial steady state, that determined the initial value for the variable e, was assumed to be one with $\bar{\imath} = 0$. The temporary policy was set to be one with $\bar{\imath} = 0.0088$, which implies a zero nominal interest rate when $n = 0$. To calculate and graph the exercises, the continuous updating procedure was approximated by the following discrete procedure:

$$e^{N+1} = (1-\gamma)e^{*,N} + \gamma e^{N} \tag{F.1}$$

where e is the whole vector of e_t. The N chosen for each figure depends of the desired n and the γ, since the approximation is given by:

$$n \approx N\gamma$$

The general algorithm for the figures can be described as:

1. **Calculate initial values**: The initial values of variables $\{y_t, \pi_t, i_t, e_{1t}, e_{2t}\}$ for all t are the ones corresponding to the steady state with $\bar{\imath} = 0$ and $\rho_t = 0$ for all t such that parameters are those in Table 1. This means that the values for all variables are zero, since all variables are defined as their deviations from that steady state. Set the initial values of the expectations $e^0_{1t} = e^0_{2t} = 0$ for all t.

2. **Introduce the change in policy**: It is one of the two following:

   - For figures 2 and 5: Maintaining the values for ϕ_y and ϕ_π as in Table 1, set $\bar{\imath} = -0.0088$ for $T = \{8, 200\}$ periods respectively, and then go back to $\bar{\imath} = 0$ forever.
   - For figures 3 and 4: Set the values $\phi_y = 0$, $\phi_\pi = 0$ and $\bar{\imath} = -0.0088$ for $T = \{8, 2000\}$ periods respectively, and then go back to the values of ϕ_y, ϕ_π of Table 1 and $\bar{\imath} = 0$ forever.

3. **Calculate the FS-PFE**: This is done by using (B.4) and (2.9).

4. **Given e_{1t}^N, e_{2t}^N, calculate y_t^N, π_t^N, i_t^N**: This is done by using (2.10) and (2.9).

5. **Given e_{1t}^N, e_{2t}^N, calculate $e_{1t}^{*,N}, e_{2t}^{*,N}$**: T do this, note that for $t \geq T$, we stay in the same steady state that we started with $y_t = \pi_t = i_t = e_{1t}^* = e_{2t}^* = 0$. Given that, calculate $e_t^{*,N}$ using equation (2.11). You can also use a recursive formulation noting that:

$$\begin{aligned} e_{1t}^{*,N} &= (1-\beta)\left(y_t^N - \frac{\sigma}{1-\beta}(\beta i_t^N - \pi_t^N)\right) + \beta e_{1t+1}^{*,N} \\ e_{2t}^{*,N} &= (1-\alpha\beta)\left(\frac{1}{1-\alpha\beta}\pi_t^N + \xi y_t^N\right) + \alpha\beta e_{2t+1}^{*,N} \end{aligned}$$

6. **Given $e_{1t}^{*,N}, e_{2t}^{*,N}$, calculate $e_{1t}^{N+1}, e_{2t}^{N+1}$**: This is done by using the formula (F.1):

$$\begin{aligned} e_{1t}^{N+1} &= (1-\gamma)e_{1t}^{*,N} + \gamma e_{1t}^N \\ e_{2t}^{N+1} &= (1-\gamma)e_{2t}^{*,N} + \gamma e_{2t}^N \end{aligned}$$

   γ is set equal to 0.02 for figures 2 and 3 and panel (a) of 5, 0.001 for figure 4 (because the approximation was inaccurate for higher values of γ) and 0 for panel (b) of figure 5.

7. **Repeat 4-6 N times**.

8. **Transform the variables to be graphed**: Use the following

$$\begin{aligned} \pi_t^{Graph} &= 100((\exp(\pi_t + \pi^*))^4 - 1) \\ y_t^{Graph} &= 100 y_t \\ i_t^{Graph} &= 100((\exp(i_t + \pi^*)/\beta).^4 - 1) \end{aligned}$$

ADDITIONAL REFERENCES

Hirsch, Morris W. and Stephen Smale (1974). *Differential Equations, Dynamical Systems, and Linear Algebra.* Academic Press.

Jury, Eliahu I. (1964). *Theory and Applications of the z-Transform Method.* Wiley.