DOCUMENTOS DE TRABAJO

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

Mariana García-Schmidt
Michael Woodford

BANCO CENTRAL DE CHILE

CENTRAL BANK OF CHILE

La serie Documentos de Trabajo es una publicación del Banco Central de Chile que divulga los trabajos de investigación económica realizados por profesionales de esta institución o encargados por ella a terceros. El objetivo de la serie es aportar al debate temas relevantes y presentar nuevos enfoques en el análisis de los mismos. La difusión de los Documentos de Trabajo sólo intenta facilitar el intercambio de ideas y dar a conocer investigaciones, con carácter preliminar, para su discusión y comentarios.

La publicación de los Documentos de Trabajo no está sujeta a la aprobación previa de los miembros del Consejo del Banco Central de Chile. Tanto el contenido de los Documentos de Trabajo como también los análisis y conclusiones que de ellos se deriven, son de exclusiva responsabilidad de su o sus autores y no reflejan necesariamente la opinión del Banco Central de Chile o de sus Consejeros.

The Working Papers series of the Central Bank of Chile disseminates economic research conducted by Central Bank staff or third parties under the sponsorship of the Bank. The purpose of the series is to contribute to the discussion of relevant issues and develop new analytical or empirical approaches in their analyses. The only aim of the Working Papers is to disseminate preliminary research for its discussion and comments.

Publication of Working Papers is not subject to previous approval by the members of the Board of the Central Bank. The views and conclusions presented in the papers are exclusively those of the author(s) and do not necessarily reflect the position of the Central Bank of Chile or of the Board members.

ARE LOW INTEREST RATES DEFLATIONARY? A PARADOX OF PERFECT-FORESIGHT ANALYSIS[*]

Mariana García-Schmidt
Banco Central de Chile

Michael Woodford
Universidad de Columbia

Abstract

A prolonged period of extremely low nominal interest rates has not resulted in high inflation. This has led to increased interest in the "Neo-Fisherian" proposition according to which low nominal interest rates may themselves cause inflation to be lower. The fact that standard models have the property that perfect foresight equilibria with a low fixed interest rate forever involve low inflation might seem to support such a view. Here, however, we argue that such a conclusion depends on a misunderstanding of the circumstances under which it makes sense to predict the effects of a monetary policy commitment by calculating the perfect foresight equilibrium (PFE). We propose an explicit cognitive process by which agents form their expectations of future endogenous variables. Under some circumstances, such as a commitment to follow a Taylor rule, a PFE can arise as a limiting case of our more general concept of reflective equilibrium. But we show that a policy of fixing the interest rate for a long period of time creates a situation in which reflective equilibrium need not resemble any PFE. In our view, this makes PFE predictions not plausible outcomes in the case of policies of the latter sort. According to our alternative approach, a commitment to maintain a low nominal interest rate for longer should always be expansionary and inflationary; but likely less so than the usual PFE analysis would imply, and much less in the case of a long-horizon commitment.

Resumen

Un largo período de tasas de interés extremadamente bajas no han provocado una alta inflación. Esto ha llevado a aumentar el interés en proposiciones "Neo-Fisherianas" en las cuales las mismas tasas bajas son las causantes de la baja inflación. Esto se ve respaldado por el hecho que modelos estándar tengan la propiedad que equilibrios de previsión perfecta con una tasa de interés baja y fija por siempre impliquen baja inflación. En este paper, en cambio, argumentamos que tal conclusión ocurre por usar equilibrios de previsión perfecta (PFE) en circunstancias no apropiadas. Proponemos un proceso cognitivo explícito con el cual los agentes forman sus expectativas de variables endógenas futuras. Bajo algunas circunstancias, tal como un compromiso de seguir una regla de Taylor, un PFE resulta como el caso límite de nuestro concepto más general de equilibrio reflectante. En cambio, una política que fija la tasa de interés por un largo período crea una situación en la que el equilibrio reflectante no se acerca a ningún PFE. Desde nuestro punto de vista, esto implica que las predicciones basadas en PFE no son plausibles. De acuerdo a nuestro enfoque alternativo, un compromiso de mantener una tasa baja y fija por mayor tiempo debería ser siempre expansivo e inflacionario, pero menos de lo implicado por el análisis PFE usual y mucho menos en el caso de un compromiso por un largo período de tiempo.

1 Perfect-Foresight Analyses of the Effects of Forward Guidance: A Paradox

One of the more notable features of recent monetary experience has been the fact that first Japan, and now more recently the U.S. as well, have gone through prolonged periods of extremely low nominal interest rates without this leading to the sort of inflationary spiral that one might have expected to follow from such a reckless experiment. Instead, inflation has remained low, below both countries' desired levels of inflation, while real activity has remained disappointing as well.

Some have proposed an interpretation of these experiences, according to which low nominal interest rates themselves may *cause* inflation to be lower. Under such a view, actually promising to keep interest rates low for a longer period than would otherwise have been expected — as both the Fed and a number of other central banks have done in the recent period[1] — would be the worst possible policy for a central bank worried that inflation will continue to run below its target, and some (beginning with Bullard, 2010, and Schmitt-Grohé and Uribe, 2010) have proposed that such a central bank should actually *raise* interest rates in order to head off the possibility of a deflationary trap. As the period over which the U.S. has kept its federal funds rate target very low has continued, views of this kind, known as "neo-Fisherian," have gained increasing currency, at least on the internet.[2]

Moreover, it might seem that even a standard textbook model of the effects of alternative monetary policy commitments would support the "neo-Fisherian" position. The most straightforward theoretical argument proceeds in two steps.[3] One first considers what should happen if a central bank were to commit to maintain the short-term nominal interest rate at an arbitrarily chosen level *forever*. According to a traditional view, famously articulated by Friedman (1968), this is not a possible experiment, because any such attempt

---

[1]See, for example, Woodford (2012) for a discussion of these experiences.

[2]See, for example, Cochrane (2015) for discussion and additional references.

[3]The argument is explained more formally in section 2.2 below.

would lead to explosive inflation dynamics that would require the central bank to abandon the policy in finite time. But in fact, many modern equilibrium models of inflation determination, including standard New Keynesian models, imply that there exist rational-expectations equilibria associated with such a policy in which inflation and other variables remain forever bounded — so that there is no reason to deny the logical possibility of the proposed thought experiment.[4] In a deterministic setting, there is typically a one-dimensional continuum of perfect foresight equilibria consistent with this policy commitment, all of which converge asymptotically to a steady state in which the constant inflation rate is the one determined by the nominal interest-rate target and the Fisher equation. This implies that inflation is higher by one percentage point for each percentage point increase in the nominal interest-rate target.

The second step in the argument notes that the outcome resulting from a given forward path for policy should not be extremely sensitive to small changes in anticipated policy that relate only to the very distant future. This means that the outcome of a commitment to keep the nominal interest rate at some level up until some finite date T should not have consequences that are very different than those that would follow from keeping the interest rate at that level forever. If keeping the interest rate low forever must eventually *lower* the inflation rate, then there must be some finite length of time such that keeping the interest rate low for that length of time *also* must eventually lower the inflation rate almost as much.

This is a paradoxical result: it seems that the very assumptions that underly common arguments for the efficacy of forward guidance — the use of a New Keynesian model and the assumption of perfect foresight (or rational expectations) — imply that a commitment to keep interest rates low for a long time should be even more disinflationary than a plan of returning sooner to a more normal policy. Yet this is not at all what standard model-based analyses of the implications of forward guidance have concluded, and it is certainly not what policymakers have assumed.

The conventional results have been challenged on another, related ground

---

[4]This is emphasized in expositions of the neo-Fisherian view such as that of Cochrane (2015).

as well. They rely upon a particular selection from among the possible PFE solutions of a New Keynesian model; but the conventional PFE selection, which is the unique bounded solution in the case of a monetary policy consistent with the Taylor principle, is not well defined when the interest rate is fixed forever. In addition, as shown by Del Negro et al. (2015), Chung (2015), and McKay et al. (2016), even in the case of a commitment to a fixed interest rate for a long but finite period, the conventionally selected equilibrium violates the principle that anticipated policy paths that differ only in the specification of policy far in the future should have similar near-term effects. Because of these reasons, it could be argued that a different equilibrium selection should be applied at least in these paradoxical cases, such as the "backward stable" equilibrium proposed by Cochrane (2016).

In this paper we consider whether a standard New Keynesian model of the effects of monetary policy requires one to accept paradoxical conclusions of the kind mentioned above.[5] We shall argue that it does not. Our quarrel, however, is not with the postulate that anticipated changes in policy sufficiently far in the future should have negligible effects on current outcomes. Rather, we deny the practical relevance of the perfect foresight solutions (or more generally, rational-expectations solutions) of the model under the experiment of a permanent or very long interest-rate peg.

Moreover, our criticism of the perfect-foresight analysis and the rejection of the Neo-Fisherian conclusion is not based on a wholesale denial of the plausibility of forward-looking expectations.[6] We believe that people are at least somewhat forward-looking; this is why central bank commitments about future monetary policy matter. Nonetheless, it may not be reasonable to expect that the outcome associated with a given policy commitment should be a PFE,

---

[5]We do not pretend to consider all of the logically possible models and policies that might be consistent with neo-Fisherian claims. For example, we do not discuss neo-Fisherian conclusions under a non-Ricardian fiscal policy as presented in Cochrane (2014); here we are solely concerned with situations in which fiscal policy is Ricardian, in a sense made precise in Woodford (2013).

[6]Friedman's view of the consequences of an interest-rate peg can be defended if one supposes that people's expectations are purely backward-looking since they will imply explosive dynamics with an interest rate peg. See Woodford (2003, sec.2.3) for discussion and references.

even when the commitment is fully credible and people have the knowledge about how the economy works that would be required for calculation of such an equilibrium.

We argue that predicting what should happen as a result of a particular policy commitment requires that one model the cognitive process by which one imagines people to arrive at particular expectations. In this paper, we offer a simple example of such an explicit model of reasoning. Under our approach, a PFE (or more generally, a rational-expectations equilibrium[7]) can be understood as a limiting case of a more general concept of reflective equilibrium, which limit may be reached under some circumstances if the process of reflection is carried far enough. Our concept of reflective equilibrium is similar to the "calculation equilibrium" proposed by Evans and Ramey (1992, 1995, 1998): we consider what economic outcomes should be if people optimize on the basis of expectations that they derive from a process of reflection about what they should expect, given both their understanding of how the economy works and their understanding of the central bank's policy intentions.

We model this process of reflection as an iterative process that adjusts the provisional forecasts that are entertained at a given stage of the process in response to the predictable discrepancy between those forecasts and what one should expect to happen if people were to behave optimally on the basis of those forecasts. Thus the process is one under which beliefs should continue to be adjusted, if the process is carried farther, unless perfect-foresight equilibrium beliefs have been reached. Like Evans and Ramey, we are interested in the theoretical question of where such a process of belief revision would end up asymptotically, but we regard it as more realistic to suppose that in practice, the process of reflection will be suspended after some finite degree of reflection, and people will act upon the beliefs obtained in this way.

The most important difference between our approach and that of Evans and Ramey is that the primary goal of their analysis is to determine how far

---

[7]We consider only deterministic environments in which, after some (possibly unexpected) change in economic fundamentals and/or the announced path of monetary policy, neither fundamentals nor policy depend on any further random events. But the reflective equilibrium defined below could also be considered in stochastic environments, in which case we could instead consider converge to a rational-expectations equilibrium.

the belief revision process should be carried forward, by specifying costs of additional calculation and a criterion for judging the benefits that should be weighed against those costs. Our concerns are instead to determine whether the process will necessarily reach a PFE even if carried forward indefinitely; to ask which PFE is reached in the case that the process converges; and to understand what determines the speed of convergence.[8]

In our view, the predictions obtained by considering the PFE consistent with a given forward path for policy are of practical relevance only in that the belief revision process converges to those PFE beliefs, sufficiently rapidly and from a large enough range of possible initial beliefs. We show below that standard conclusions about equilibrium determination under a Taylor rule (when the zero lower bound does not constrain policy) can be justified in this way; our analysis not only provides a reason for interest in the perfect-foresight predictions, but explains why one particular PFE solution should be regarded as the relevant prediction of the model.[9]

If, instead, a particular perfect foresight equilibrium cannot be reached under the belief revision process, except by starting from extremely special initial beliefs, then we do not think it is plausible to expect actual outcomes to resemble those PFE outcomes.[10] And if the belief revision process does not converge, or if it converges only very slowly, then we think that the beliefs and economic outcomes that should be observed in practice need have little to do with the set of perfect foresight equilibria. We show below that the thought experiment of an interest-rate peg that is maintained forever produces a situation of this kind: while perfect foresight equilibria do indeed exist, the belief revision process that we consider does not converge to any of them, and the set of reflective equilibria resulting from different finite degrees of reflection do not resemble perfect foresight equilibria. It is important to note however, that even in cases of slow or no convergence, it may well be possible to

---

[8]For a complete discussion of related proposals, refer to section 2.4 of García-Schmidt and Woodford (2015).

[9]This also addresses the critique in Cochrane (2011) of the standard New Keynesian literature.

[10]This is our view of the bounded perfect foresight equilibria when the interest rate is kept fixed forever and of the "backward stable" PFE solutions analyzed by Cochrane (2016) in the case of a temporary interest-rate peg.

derive *qualitative* conclusions about the effects on a reflective equilibrium that should be expected from changing policy in a particular direction; and these may differ, even as to sign, from those that would be suggested by considering the set of perfect foresight equilibria.

In particular, we show that in our model, a commitment to maintain a low nominal interest rate for any fixed length of time will typically result in increased aggregate demand, increasing both output and inflation in the near term, though the exact degree of stimulus that should result depends (considerably) on the assumed degree of reflection. This is true even in the limit of a perpetual interest-rate peg. Thus consideration of the reflective equilibrium resulting from a finite degree of reflection yields conventional conclusions about the sign of the effects of commitments to lower interest rates in the future, and does so without implying any non-negligible effects of changing the specification of policy only very far in the future. Hence the reflective equilibrium analysis avoids both of the paradoxical conclusions that a PFE analysis requires one to choose between: affirming either that maintaining low nominal interest rates must eventually be deflationary, or that the outcome implied by a given policy commitment can depend critically on the specification of policy extremely far in the future.

We proceed as follows. Section 2 introduces our New Keynesian model of inflation and output determination under alternative monetary policies and alternative assumptions about private-sector expectations, and the belief revision process that underlies our proposed concept of reflective equilibrium. Section 3 considers reflective equilibrium when the forward path of monetary policy is specified by a Taylor rule, and both the path of policy and the economy's exogenous fundamentals are such that the ZLB never binds in equilibrium. We allow the Taylor rule to involve a time-varying intercept (or inflation target), to analyze a type of forward guidance. By contrast, section 4 considers reflective equilibrium in the less well-behaved case of an expectation that the short-term interest rate will remain fixed until some horizon T, and then revert to a Taylor rule thereafter; and also the limiting case of a commitment to keep the interest rate fixed forever. Section 5 offers concluding reflections.

2 Reflective Equilibrium in a New Keynesian Model

We expound our concept of reflective equilibrium in the context of a log-linearized New Keynesian (NK) 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 when policy is temporarily constrained by the zero lower bound (e.g., Eggertsson and Woodford, 2003; Werning, 2012; McKay et al., 2016; Cochrane, 2016).[11] As in the analyses of Evans and Ramey (1992, 1995, 1998), we must begin by specifying the *temporary equilibrium relations* that map arbitrary subjective expectations about future economic conditions into market outcomes; these relations play a crucial role in the process of reflection that we wish to model, in addition to being required in order to predict what should happen if people's beliefs do not converge to PFE beliefs. The presentation here largely follows Woodford (2013), where the derivations are discussed in more detail.

2.1 Temporary Equilibrium Relations

The economy is made up of identical, infinite-lived households. Each household i seeks to maximize a discounted flow of utility

$$\hat{E}_t^i \sum_{T=t}^{\infty} \exp[-\sum_{s=t}^{T-1} \rho_s][u(C_T^i) - v(H_T^i)] \tag{2.1}$$

when planning their path of consumption, looking forward from date t. Here C_t^i is a Dixit-Stiglitz aggregate of the household's purchases of differentiated consumer goods, H_t^i is hours worked, the sub-utility functions satisfy $u' > 0, u'' < 0, v' > 0, v'' \geq 0$, and ρ_t is a possibly time-varying discount rate. We allow for a non-uniform discount rate to introduce a reason why the ZLB may temporarily constrain monetary policy; the fact that intra-temporal

---

[11]Werning (2012) and Cochrane (2016) analyze a continuous-time version of the model, but the structure of their models is otherwise the same as the discrete-time model considered here.

preferences are uniform over time will allow the efficient level of output to be constant over time.[12] The operator $\hat{E}_t^i$ indicates that this objective is evaluated using the future paths of the variables implied by the household's subjective expectations, which need neither be model-consistent nor common across all households.

For simplicity, there is assumed to be a single traded asset each period: one-period riskless nominal debt (a market for which must exist in order for the central bank to control a short-term nominal interest rate). Each household also owns an equal share of each of the firms, but these shares are assumed not to be tradeable. In the present exposition, we abstract from fiscal policy, by assuming that there are no government purchases, government debt, or taxes and transfers.[13]

We can then solve for the household's optimal expenditure plan, $\{C_T\}$ for dates $T \geq t$, as a function of the expected paths of real income, the one-period nominal interest rate, and the rate of inflation. We log-linearize the optimal decision rule around the constant plan that is optimal in the event that $\rho_T = \bar{\rho} > 0$ for all $T \geq t$, the inflation rate is expected to equal the central bank's target rate π^* in all periods, and real income and the nominal interest rate are also expected to be constant in all periods at values that represent a PFE for a monetary policy that achieves the inflation target at all times.[14] We obtain

$$c_t^i \; = \; (1-\beta)\hat{b}_t^i + \sum_{T=t}^{\infty} \beta^{T-t}\, \hat{E}_t^i \left\{ (1-\beta)y_T \; - \; \beta\sigma\left(i_T - \pi_{T+1} - \rho_T\right) \right\} \qquad (2.2)$$

where $\{y_T, i_T, \pi_T\}$ are the expected paths of real income (or aggregate output, in units of the Dixit-Stiglitz aggregate), the nominal interest rate, and inflation, and all variables appearing in the equation are measured as log devia-

---

[12]In Woodford (2013), a more general version of the model is presented, in which a variety of other types of exogenous disturbances are allowed for.

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

[14]We assume that $\pi^* > -\bar{\rho}$, so that the nominal interest rate in this steady state is positive.

tions from their steady-state values. Here $\beta \equiv \exp[-\bar{\rho}] < 1$ is the steady-state discount factor and $\sigma > 0$ is the intertemporal elasticity of substitution of consumer expenditure. Note that (2.2) generalizes the familiar "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 can correctly forecast the variation over time (if any) in their discount rate, so that $\hat{E}_t^i \rho_T = \rho_T$ for all $T \geq t$.[15] The subjective expectations regarding future conditions that matter for a household's expenditure decision can be collected in a single expectational term, allowing us to rewrite (2.2) as

$$c_t^i = (1 - \beta)y_t - \beta\sigma i_t + \beta g_t + \beta \hat{E}_t^i v_{t+1}^i, \qquad (2.3)$$

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

$$v_t^i \equiv \sum_{T=t}^{\infty} \beta^{T-t} \hat{E}_t^i \{(1 - \beta)y_T - \sigma(\beta i_T - \pi_T)\}$$

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_t^i over households i, the individual decision rules (2.3) aggregate to an *aggregate demand (AD) relation*

$$y_t = g_t - \sigma i_t + e_{1t}, \qquad (2.4)$$

where

$$e_{1t} \equiv \int \hat{E}_t^i v_{t+1}^i \, di$$

---

[15]This means that expectations regarding future preference shocks are treated differently in (2.3) than in the expression given in Woodford (2013). The definition of the composite expectational variable, v_t^i, is correspondingly different.

is a measure of average subjective expectations.

The continuum of differentiated goods are produced by Dixit-Stiglitz monopolistic competitors, who each adjust their prices only intermittently; as in the Calvo-Yun model of staggered pricing, only 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 using the assumption that prices that are not reconsidered in any given period are automatically increased at the target rate π^*.[16] If a firm j reconsiders its price in period t, it chooses a price that it expects to maximize the present discounted value of profits in all future states prior to the next reconsideration of its price, given its subjective expectations regarding the evolution of aggregate demand $\{y_T\}$ for the composite good 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

$$p_t^{*j} \;\; = \;\; (1 - \alpha\beta) \sum_{T=t}^{\infty} (\alpha\beta)^{T-t} \hat{E}_t^j [p_T + \xi y_T - \pi^*(T - t)] - (p_{t-1} + \pi^*) \quad (2.5)$$

where p_t^{*j} is the amount by which $j's$ log price exceeds the average of the prices that are not reconsidered, $p_{t-1} + \pi^*$, $\xi > 0$ measures the elasticity of a firm's optimal relative price with respect to aggregate demand.[17] The operator $\hat{E}^j[\cdot]$ indicates that what matter are the subjective expectations of firm j regarding future market conditions.

Again, the terms on the right-hand side of (2.5) involving subjective expectations at various horizons can be collected in a single composite term, $\alpha\beta \hat{E}_t^j p_{t+1}^{*j}$. Aggregating across the prices chosen in period t, we obtain an *aggregate supply (AS) relation*

$$\pi_t \;\; = \;\; \kappa y_t \; + \; (1 - \alpha)\beta \, e_{2t} \quad (2.6)$$

---

[16]This allows us to assume a positive steady-state inflation rate — which is important for the quantitative realism of the numerical examples below — while at the same time retaining the convenience of a steady state in which the prices of all goods are identical.

[17]The parameter ξ is thus a measure of the degree of "real rigidities." See Woodford (2003, chap. 3) for a detailed discussion of its dependence on underlying parameters.

where $\pi_t \equiv p_t - p_{t-1} - \pi^*$ is inflation in excess of the target rate,

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

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

$$i_t = \bar{\imath}_t + \phi_\pi \pi_t + \phi_y y_t \tag{2.7}$$

where the response coefficients satisfy $\phi_\pi, \phi_y \geq 0$. We allow for a possibly time-varying intercept in order to consider the effects of announcing a transitory departure from the central bank's normal reaction function. Equations (2.4), (2.6) and (2.7) then comprise a three-equation system, that determines the temporary equilibrium (TE) 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 expectations (e_{1t}, e_{2t}). Under our sign assumptions, TE values are uniquely determined, linear functions of the vector of disturbances and the vector of subjective expectations (see the Appendix for details).

It is useful to note the relationship between subjective expectations and the actual values of the variables that people seek to forecast. The two sufficient statistics for subjective expectations e_{it}, for $i = 1, 2$, can be written as:

$$e_{it} = (1 - \delta_i) \sum_{j=0}^{\infty} \delta_i^j \bar{E}_t a_{i,t+j+1}, \tag{2.8}$$

where

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

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 the operator $\bar{E}_t[\cdot]$ indicates the average of the population's forecasts at date t.[18]

We can then use the TE relations to solve for the equilibrium values of the variables a_{it} that people seek to forecast as linear functions of the current vector of disturbances and current average expectations. This solution can be written in the form

$$
a_t = M e_t + m \omega_t, \tag{2.9}
$$

where a_t is the vector (a_{1t}, a_{2t}), e_t is the vector (e_{1t}, e_{2t}), ω_t is the vector $(g_t, \bar{\imath}_t)$, and the matrices of coefficients are given in the Appendix. The system (2.9) shows how expectations determine the endogenous variables that are themselves being forecasted in those expectations, as indicated by (2.8).

2.2 Perfect Foresight Equilibrium

The assumption of perfect foresight equilibrium adds to the above model the further assumption that the expected paths for output, inflation and the interest rate (and hence the expected paths for the variables $\{a_t\}$) are precisely the paths for those variables implied by the TE relations under those expectations. Thus a PFE corresponds to sequences $\{a_t, e_t\}$ that both satisfy (2.9) each period and satisfy (2.8) when the equilibrium paths $\{a_t\}$ are substituted for the average expectations in those equations.

It can be shown (see Woodford, 2013) that under the PFE assumption, the TE relations (2.4) and (2.6) imply that the paths of output, inflation and the interest rate must satisfy difference equations of the form

$$
y_t = y_{t+1} - \sigma(i_t - \pi_{t+1} - \rho_t) \tag{2.10}
$$

---

[18]While we still allow for the possibility of heterogeneous forecasts, from here on we assume that the distribution of forecasts across households is the same as across firms.

$$\pi_t = \kappa y_t + \beta \pi_{t+1} \tag{2.11}$$

which are simply perfect-foresight versions of the usual "New Keynesian IS curve" and "New Keynesian Phillips curve" respectively. Using the policy specification (2.7) to eliminate i_t, one obtains a pair of difference equations that can be written in the form

$$x_t = B\,x_{t+1} + b\,(\rho_t - \bar{\imath}_t) \tag{2.12}$$

where x_t is the vector consisting of (y_t, π_t), and the matrix B and vector b are defined in the Appendix.

Under our sign assumptions for the model coefficients, the matrix B is invertible, and the system (2.12) can be uniquely solved for x_{t+1} as a function of x_t and the period t disturbances. One then obtains a two-parameter family of possible PFE solutions consistent with any given forward paths for the disturbances, corresponding to the set of possible choices for the elements of x_0. The asymptotic behavior of these solutions as t is made large depends as usual on the eigenvalues of the matrix B.

As shown in the Appendix, the matrix B has both eigenvalues inside the unit circle if and only if

$$\phi_\pi + \frac{1-\beta}{\kappa}\phi_y > 1 \tag{2.13}$$

so that the "Taylor Principle" is satisfied. In this case, there is a unique bounded PFE solution for the sequences $\{x_t\}$ corresponding to any bounded sequences $\{\rho_t, \bar{\imath}_t\}$, obtained by "solving forward" the system (2.12) to obtain

$$x_t = \sum_{j=0}^{\infty} B^j b\,(\rho_{t+j} - \bar{\imath}_{t+j}). \tag{2.14}$$

When this is uniquely defined, we shall call this the "forward stable" PFE (FS-PFE). It is common to regard this as the relevant prediction of the model in such a case;[19] below we shall provide a justification for this in terms of our concept of reflective equilibrium.

This solution implies that in the case of a sufficiently transitory change in

---

[19]See however Cochrane (2011) for objections to this interpretation.

policy, a reduction of $\bar{\imath}_t$ (for a given path of the real disturbance) must be both expansionary and inflationary, while the nominal interest rate is temporarily reduced (though by less than the reduction in $\bar{\imath}_t$). In the case of a sufficiently persistent shift in $\bar{\imath}_t$, output, inflation and the nominal interest rate are *all* predicted to increase, because of the endogenous effect of the output and inflation increases on the central bank's interest-rate target;[20] but even in this case, a downward shift in the reaction function (reducing the interest-rate target implied by any given current levels of inflation and output) is inflationary rather than deflationary.

If the inequality in (2.13) is reversed, the matrix B instead has two real eigenvalues satisfying

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

so that the larger is outside the unit circle. In particular, this is true if the central bank fixes the forward path for the nominal interest rate (the case $\phi_\pi = \phi_y = 0$), regardless of whether this path is constant. In this case, there is no longer a unique bounded solution; instead, assuming again that the sequences $\{\rho_t, \bar{\imath}_t\}$ are bounded, there is a bounded PFE solution

$$x_t = v_1(e_1'b) \sum_{j=0}^{\infty} \mu_1^j \left(\rho_{t+j} - \bar{\imath}_{t+j}\right) \ - \ v_2(e_2'b) \sum_{j=1}^{t} \mu_2^{-j} \left(\rho_{t-j} - \bar{\imath}_{t-j}\right) \ + \ \chi \, v_2 \mu_2^{-t}$$

(2.15)

in the case of any real number χ.[21] Since such solutions necessarily exist, the PFE analysis gives us no reason to suppose that there is anything problematic about a commitment to fix a path for the nominal interest rate, including a commitment to fix it at a constant rate forever.

Now suppose that not only are the exogenous disturbance sequences bounded, but that after some finite date T, they are expected to be constant: $\rho_t = \rho_{LR}$

---

[20]In a more realistic model than the simple NK model used in this paper, there will be delays in the effect of the policy change on output and inflation. It is then possible to have an initial decline in the nominal interest rate in the case of an expansionary monetary policy shock, even in the case of a relatively persistent shift in the central-bank reaction function, as shown in Woodford (2003, secs. 5.1-5.2).

[21]In this expression, v_i is the right eigenvector corresponding to eigenvalue μ_i, e_i' is the left eigenvector corresponding to that same eigenvalue, and we normalize the eigenvectors so that $e_i'v_i = 1$ for each i.

and $\bar{\imath}_t = \bar{\imath}_{LR}$ for all $t \geq T$, where the long-run values need not equal zero. We show in the Appendix that in *any* of the continuum of bounded PFE solutions, the elements of x_t converge asymptotically to long-run values

$$\pi_{LR} = \bar{\imath}_{LR} - \rho_{LR}, \qquad y_{LR} = \frac{1-\beta}{\kappa}(\bar{\imath}_{LR} - \rho_{LR}). \qquad (2.16)$$

One observes that the long-run inflation rate increases one-for-one with increases in the long-run interest-rate target. Hence if we suppose that the economy must follow one or another of the PFE associated with the central bank's policy commitment, we would conclude that a lower path for the nominal interest rate must at least eventually result in a lower rate of inflation.

One might obtain even stronger conclusions under further assumptions about how to select a particular solution from among the set of PFE. Consider the simple case of a policy commitment under which the interest rate will be fixed at one level $\bar{\imath}_{SR}$ for all $0 \leq t < T$, and another (possibly different) level $\bar{\imath}_{LR}$ for all $t \geq T$, and let us suppose that $\rho_t = 0$ for all t.[22]

Assuming further (as a sort of "minimum-state-variable" criterion) that the solution should depend only on $t - T$ rather than on the absolute magnitude of either t or T, the perfect foresight analysis yields a "neo-Fisherian" conclusion, since the inflation rate can be expressed as

$$\pi_t = \lambda(t - T)\bar{\imath}_{SR} + (1 - \lambda(t - T))\bar{\imath}_{LR},$$

where the sequence of weights $\{\lambda(t - T)\}$ (all between zero and 1) depend only on the distance in time from the date of the policy shift. Increasing both $\bar{\imath}_{SR}$ and $\bar{\imath}_{LR}$ by the same number of percentage points is predicted to increase the inflation rate *in all periods* by exactly that same number of percentage points. Increasing only *one* of the interest rates is also predicted to increase the inflation rate both initially and later.

The conclusions that one obtains about the sign of the effects of a shift in the anticipated path of $\{\bar{\imath}_t\}$ on inflation seem then to depend crucially on the

---

[22]Given the linearity of the model, it is reasonable to suppose that the prediction in the case of any disturbance sequences $\{\rho_t, \bar{\imath}_t\}$ can be expressed as the sum of a predicted effect of $\{\rho_t\}$ and that of $\{\bar{\imath}_t\}$

magnitude of the reaction coefficients (ϕ_π, ϕ_y), if one believes the results of the perfect foresight analysis. We shall argue however, that the conclusions of PFE analysis are misleading in the case just discussed, in which the "Taylor principle" is violated. This leads us to consider whether the PFE paths just discussed are ones that can be justified as resulting from beliefs that people would arrive at under a process of reflection.

2.3 Reflective Equilibrium

Why should people have the particular expectations about the future that are assumed in a perfect foresight equilibrium? One answer could be that, if they expected a future path for the economy of any *other* type, action on the basis of their expectations would produce outcomes that disconfirm those expectations. One might suppose, then, that experience should sooner or later disabuse people of any expectations that are not consistent with a PFE. And if one supposes that people have sufficient structural knowledge (including understanding of the central bank's intentions regarding future policy), one might even think that they should be able to recognize the inconsistency between their expectations and what they should expect to happen if others were also to think that way — allowing them to refine their beliefs on the basis of their understanding of how the economy works, even prior to any experience with the current economic disturbance or policy regime. Here we explore the conditions under which a PFE might arise (or under which outcomes would at least approximate a PFE) as a result of reflection of the latter sort.

We model a process of reflection by a decisionmaker (DM) who understands how the economy works — that is, who knows the TE relations (2.4) and (2.6) — and who also understands the policy intentions of the central bank, in the sense of knowing the policy rule (2.7) that will determine policy in all future periods. However, while the DM understands (and fully believes) the announcement of what the central bank will do, she does not know, without further reflection, what this implies about the future evolution of national income, inflation, or the resulting level of interest rates (unless the policy rule specifies a fixed interest rate).

16

The assumed structural knowledge can however be used to refine her expectations about the evolution of those variables. Suppose that the DM starts with some conjecture about the future evolution of the economy, which we can summarize by paths for the variables $\{e_t\}$ for each of the dates $t \geq 0$, where $t = 0$ means the date at which the economy's future evolution is being contemplated. She can then ask: suppose that others were sophisticated enough to have exactly these expectations (on average), both now and at all of the future dates under consideration. What path for the economy should *she* expect, given her structural knowledge, under this conjecture about others' average expectations?

Under such a conjecture $\{e_t\}$, the TE relations imply unique paths for the variables $\{a_t\}$, given by (2.9). From these predictions the DM can infer implied paths $\{e_t^*\}$ for all $t \geq 0$, where for each date t, e_t^* are the forecasts that would be *correct* at that date if the economy evolves in the way implied by the TE relations for $\{e_t\}$. This deduction yields an affine operator

$$e^* = \Psi e$$

mapping sequences $\{e_t\}$ of conjectured expectations into sequences $\{e_t^*\}$ of correct forecasts of the same variables.[23] Note that the operator Ψ is purely forward-looking; in fact, we can write

$$e_t^* = \sum_{j=1}^{\infty} \psi_j e_{t+j} + \sum_{j=1}^{\infty} \varphi_j \omega_{t+j} \tag{2.17}$$

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

$$\psi_j \equiv (I - \Lambda)\Lambda^{j-1}M, \qquad \varphi_j \equiv (I - \Lambda)\Lambda^{j-1}m, \qquad \Lambda \equiv \begin{pmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{pmatrix}$$

for all $j \geq 1$.

---

[23]Note that the definition of the operator Ψ depends on the sequences of fundamental perturbations $\{\omega_t\}$. To simplify notation, we suppress these additional arguments. We shall be interested in the application of this operator to different possible conjectured beliefs $\{e_t\}$, holding fixed the fundamentals.

We suppose that an awareness that the implied correct sequences e^* differ from the conjectured sequences e should constitute a reason to doubt the reasonableness of expecting people to hold the conjectured beliefs. But what should one expect instead? We propose that the conjectured beliefs should be adjusted in the direction of the discrepancy between the model prediction on the basis of the conjectured beliefs and the conjectured beliefs themselves. Specifically, we consider a process of belief revision described by a differential equation

$$\dot{e}_t(n) \; = \; e_t^*(n) \; - \; e_t(n), \qquad\qquad (2.18)$$

where the continuous variable $n \geq 0$ indexes how far the process of reflection has been carried forward, $e_t(n)$ is the conjecture regarding average beliefs in period t at stage n of the process, $e_t^*(n)$ is the implied correct forecast in period t if average expectations are given by the stage n conjectures, and $\dot{e}_t(n)$ is the derivative of $e_t(n)$ with respect to n.[24]

We suppose that the process of reflection starts from some initial "naive" conjecture about average expectations $e_t(0)$, and that the differential equations (2.18) are then integrated forward from those initial conditions. This initial conjecture might be based on the forecasts that *would* have been correct, but for the occurrence of the unusual shock and/or the change in policy that are the occasion for the process of reflection about what to expect in light of new circumstances. (One might suppose that but for these changes, the situation would have been a sufficiently routine one for people to have learned how to accurately forecast the economy's evolution.) 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 the sequences $\{e_t^*(n)\}$ still differ from $\{e_t(n)\}$.

By a *reflective equilibrium* we mean a situation in which output, inflation and the nominal interest rate at some date (here numbered date 0) are determined by the TE relations, using the average subjective expectations $e_0(n)$ at the stage n at which the belief revision process is terminated.[25] We may also

---

[24]In García-Schmidt and Woodford (2015, sec. 2.3) we discuss possible interpretations of the dynamics specified by (2.18).

[25]The reflective equilibrium of degree n can already be viewed as one in which there is

refer to the entire sequence of outcomes in periods $t \geq 0$ implied by the TE relations if average subjective expectations in each period are given by $e_t(n)$ as representing a reflective equilibrium of degree n. One should however only expect this entire sequence of outcomes to be realized on the assumption that the *same process of reflection* would determine beliefs and hence actions in each of the subsequent periods. This would make sense only if one supposes that the assumptions used as inputs to the process of reflection do not change in later periods in the light of additional observations, or that the process of reflection is only undertaken once. So, here we consider only the beliefs resulting from a one-time process of reflection, and how similar or not these should be to PFE beliefs.

The proposed concept of reflective equilibrium will not generally lead to a unique prediction as to how the economy should evolve as a result of a specific policy commitment; the reflective equilibrium outcome will depend 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 (2.18) converge globally (or at least for a large enough set of possible initial conditions) to a particular PFE, and furthermore converge rapidly enough, then a quite specific prediction will be possible under fairly robust assumptions. This is the case in which it would be justifiable to use a PFE (the *specific* PFE that represents this limit of the process of belief revision) as a prediction for what should happen under the policy commitment in question. We show below that this can justify use of the FS-PFE when policy conforms to a Taylor rule.

And even when the reflective equilibrium does not provide a precise quantitative prediction, but only indicates a range of possibilities, such a qualitative prediction may nonetheless be of interest. In some cases, *all* of the possible outcomes allowed by this concept may be quite different from any of the possible PFE. We show that this is true in our model in the case of a commitment to a fixed interest rate for a long period of time.

------

a distribution of different levels of reflection across the population, with n indicating the mean level of reflection rather than a level common to all individuals. See García-Schmidt and Woodford (2015) for further discussion.

3   Convergence of Reflective Equilibrium to Perfect Foresight: The Case of a Taylor Rule

Here we show that under some circumstances, the PFE analysis (with a correct equilibrium selection) can be justified as an approximation to a reflective equilibrium, and that (for some parameter values) the degree of reflection required to approach the PFE outcome need not even be too large. The case that we consider is that in which monetary policy is expected to be specified by a Taylor-type rule of the form (2.7), where the response coefficients ϕ_π, ϕ_y are assumed to be constant over time, though we allow a time-varying path for the intercept $\{\bar{\imath}_t\}$. The allowance for time-variation in the intercept allows us to analyze policy experiments in which there may be a commitment to conduct a "looser" policy for a specified period of time (temporary reduction of the intercept), before returning to the central bank's normal reaction function. We further assume that these endogenous responses satisfy (2.13), so that there is a unique bounded PFE solution in the case of any bounded sequences $\{g_t, \bar{\imath}_t\}$, given by (2.14).[26]

We assume in this section that under the reflective equilibrium dynamics, the zero lower bound never binds, so that it is in fact feasible for the central bank's interest-rate target to satisfy (2.7) 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 for the interest rate implied by the TE relations to always be non-negative. This section applies as long as the shock to the economy, the policy response to it, and any associated shift in the initially conjectured expectations are all small enough departures from the long-run steady state. We defer until the next section consideration of the case in which a large shock causes the ZLB to constrain policy for some period of time.

---

[26]Note that any bounded sequence $\{g_t\}$ uniquely determines a bounded sequence $\{\rho_t\}$, given by $\rho_t = \sigma^{-1}[g_t - \beta g_{t+1}]$.

3.1 Exponentially Convergent Belief Sequences

Our results on the convergence of reflective equilibrium as the degree of re-flection 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 $\{x_t\}$ 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

$$x_t = x_\infty + \sum_{k=1}^{K} a_k \lambda_k^{t-\bar{T}}, \tag{3.1}$$

where x_∞ and the $\{a_k\}$ are a finite collection of real coefficients, and the $\{\lambda_k\}$ are real numbers satisfying $|\lambda_k| < 1$. This places no restrictions on the behavior of 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 (elements of the vector).

We shall consider only the case in which the initial belief sequence $\{e_t(0)\}$ converges exponentially. Note that the TE relations (2.8)–(2.9) imply that if the sequence of fundamentals $\{\omega_t\}$ converges exponentially, and a conjecture $\{e_t\}$ regarding average subjective expectations converges exponentially as well, then the correct expectations $\{e_t^*\}$ implied by this conjecture also converge exponentially. Thus the operator Ψ maps exponentially convergent belief sequences into exponentially converging belief sequences.

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 (2.18) proceeds (that is, as n increases), assuming that fundamentals $\{\omega_t\}$ and the initial "naive" conjecture $\{e_t(0)\}$ converge exponentially.[27] In this case we have our first proposition:

Proposition 1 *Consider the case of a shock sequence $\{g_t\}$ that converges ex-*

---

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

ponentially, and let the forward path of policy be specified by a sequence of reaction functions (2.7), where the coefficients (ϕ_π, ϕ_x) are constant over time and satisfy (2.13), 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 (2.18) 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 $t \geq 0$.

Further details of the proof are given in the Appendix.

This result has several implications. First, it shows how a PFE can arise through a process of reflection of the kind proposed in section 2.3. But further, it indicates that only one of the two-dimensional continuum of solutions to the difference equations (2.12) represents a PFE that can be reached in this way.[28] Thus it provides a justification for selecting the FS-PFE as the relevant perfect-foresight prediction of the model.

Proposition 1 also shows that the proposal to use reflective equilibrium need not mean that one cannot obtain predictions of any precision. In the case considered here, the reflective equilibrium predictions are quite similar, for all sufficiently large values of n. They are also similar regardless of the initial conjecture that is assumed, as long as the initial conjecture is not extremely distant from the beliefs associated with the long-run steady state. Finally, we see that the FS-PFE provides a useful approximation of the reflective equilibrium. The accuracy of this approximation should be greater the greater the degree of reflection that one assumes.

---

[28]If we accept the reasonableness of starting from an initial conjecture that is well-behaved in the sense assumed in the proposition. This includes, for example, the "naive" hypothesis that people's expectations should be unaffected by either the shock that has occurred or the resulting change in policy.

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 degree of reflection may not be implausibly large. We illustrate this by considering a numerical example.

Figure 1 considers an experiment in which the intercept $\bar{\imath}_t$ is lowered for 8 quarters (periods $t = 0$ through 7 of the quarterly model), but is expected to return to its normal level from quarter 8 onward. 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 $\phi_\pi = 1.5, \phi_y = 0.5/4$.[29] 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 $\xi = 0.125$.[30] Among other things, these imply a long-run steady-state value for the nominal interest rate of 3.23 percent per annum.[31]

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 the reflective equilibria associated with any $n \geq 0$.[32] In computing the reflective equilibria shown in Figure 1, we assume an initial "naive" conjecture under which expectations continue to be those that are correct in the steady state with 2 percent inflation. Finally, for simplicity we consider only a pure

---

[29]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 (2.7) are quarterly rates rather than the annual rates used in Taylor's formula.

[30]We do not pretend to offer a quantitatively realistic analysis of alternative policies that should have been available during the Great Recession; our goal in this paper is purely to explicate the conditions under which perfect foresight analysis of monetary policy commitments makes a greater or lesser amount of sense. The parameter values proposed by 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 a rational-expectations analysis.

[31]This means that the intercept of the central-bank reaction function assumed in the long run is smaller here than in Taylor (1993); we assume the value that (in our model) is consistent with achievement of the 2 percent inflation target in the long-run steady state.

[32]As shown in Figure 1, 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 the belief-revision dynamics do not change expectations regarding any of the periods from $t = 8$ onward.
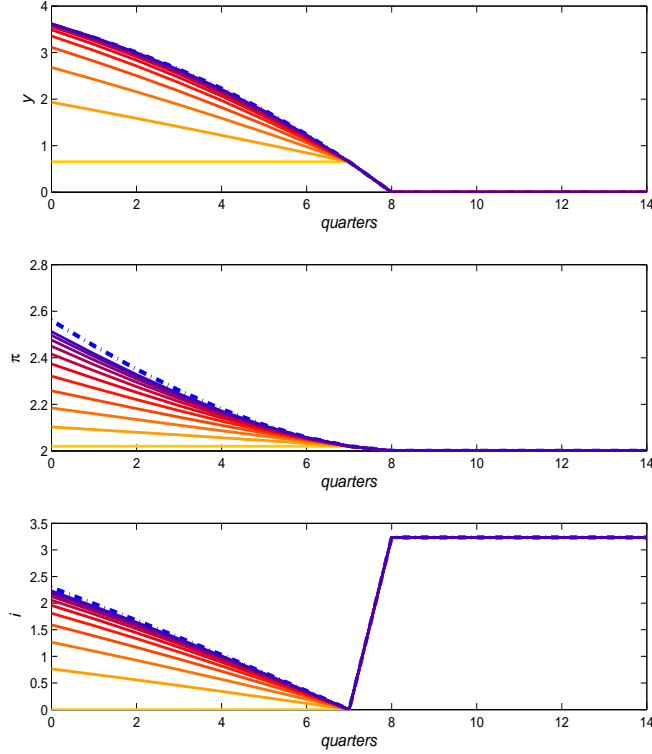
Figure 1: Reflective equilibrium outcomes for $n = 0$ through 4 (progressively darker lines) compared with the FS-PFE solution (dash-dotted line), when the Taylor-rule intercept is reduced for 8 quarters.

temporary loosening of monetary policy, not motivated by any real disturbance (so that $g_t = 0$ for all t).[33]

The three panels of the figure show the TE paths of output, inflation and the nominal interest rate,[34] in reflective equilibria corresponding to successively

---

[33]Because our model is linear, we can separately compute the perturbations of all variables implied by a pure monetary policy shift, the perturbations implied by a real disturbance (assuming no change in monetary policy and no change in the initial conjecture), and the perturbations implied 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 a shift in the initial conjecture. In the figure, we isolate the pure effect of an announced loosening of monetary policy, to last for a known length of time.

[34]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

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 $n = 0$ paths do not represent the naive beliefs, but rather the paths that it would be correct to expect, if people on average hold the naive beliefs. The relatively aggressive 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 PFE. 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 TE levels of output and inflation in the earlier quarters increase (and the nominal interest rate increases as well, through an endogenous policy reaction). This effect is greater the larger the number of future quarters about which expectations of output and inflation are revised upward.

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 from the figure 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, for this kind of policy experiment. Already when $n = 2$, the predicted reflective equilibrium responses for both output and inflation differ

---

π_t and i_t are reported as annualized rates, and the units are again percentage points; thus "2" means two percent per annum.

from the PFE responses by less than 10 percent in any quarter. This means that if the average member of the population is expected to be capable of iterating the Ψ mapping at least twice,[35] one should predict outcomes approximately the size of the PFE outcomes. When $n = 4$, the reflective equilibrium output responses differ from the PFE 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.

This provides a good example of the kind of situation in which, in our view, a perfect foresight equilibrium analysis of the effects of a monetary policy commitment can make sense. Note that is specifically the FS-PFE, rather than any other solution to the difference equations (2.12), that provides a good approximation to a reflective equilibrium.

3.3 Effects of a Policy Change Far in the Future

The paradox posed in section 1 involves arguments about the effects of an expectation that policy will be changed *permanently,* rather than for only a few quarters as in Figure 1, and questions about how much it can matter what is assumed about policy extremely far in the future. Here we consider these issues in the case of the class of policies discussed above.

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 $t \geq T$. (The policy experiment considered in Figure 1 is one example of a policy in this class, with $\bar{\imath}_{SR} < 0$, $\bar{\imath}_{LR} = 0$, and $T = 8$. We again assume that both $\bar{\imath}_{SR}$ and $\bar{\imath}_{LR}$ are high enough that the ZLB never binds.) How does the effect of such a policy commitment vary depending on the choice of the horizon T? In particular, should the effect be similar for

---

[35]Associated experimental evidence find that a value around 2 is appropriate. Arad and Rubinstein (2012) find that their subjects have a mean "level of thinking" of 2.2. Camerer et al. (2004), however, conclude that "an average of 1.5 steps [of iterated best response] fits data from many games" [p. 861]. These experimental results relate to subjects' play in one-shot games, where the strategic considerations have been explained to the players, but they have no *experience*. One might expect that the realistic mean degree of reflection n will be higher in cases where people have some degree of prior experience with the policy regime in question.

all large enough values of T?

As in the case considered in Figure 1, 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. Because our model is purely forward-looking, and $\bar{\imath}_t, g_t$, and $e_t(0)$ are each the same for all $t \geq T$, the belief-revision dynamics (2.18) result in $e_t(n)$ having the same value for all $t \geq T$. Let this value be denoted $e_{LR}(n)$. We see that it must evolve according to

$$\dot{e}_{LR} = [M - I]\,e_{LR} + m_2\,\bar{\imath}_{LR} \tag{3.2}$$

starting from the initial condition $e_{LR}(0) = 0$, where m_2 is the second column of the matrix m in (2.9).

Let us suppose that the quantity on the left-hand side of (2.13) is not exactly equal to 1;[36] in this case we can show (see the Appendix) that $M - I$ is non-singular. The solution to (3.2) is then easily seen to be[37]

$$e_{LR}(n) = [I - \exp[n(M - I)]]\,\bar{e}_{LR}^{PF} \tag{3.3}$$

for all $n \geq 0$, where

$$\bar{e}_{LR}^{PF} \equiv [I - M]^{-1}\,m_2\,\bar{\imath}_{LR} \tag{3.4}$$

is the unique rest point of the dynamics (3.2).

Note that $\bar{e}_{LR}^{PF}$ is also the stationary vector of average expectations associated with the unique PFE steady state. If the reaction coefficients (ϕ_π, ϕ_y) satisfy the Taylor Principle (2.13), then (as shown in the Appendix) both eigenvalues of $M - I$ have negative real part, and

$$\lim_{n \to \infty} \exp[n(M - I)] = 0. \tag{3.5}$$

---

[36]This condition is satisfied by generic reaction functions of the form (2.7) 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.

[37]See, for example, Hirsch and Smale (1974, p. 90).

It then follows from (3.3) that

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

so that the reflective equilibrium in any period $t \geq T$ converges to the PFE steady state associated with the long-run policy (which is also the FS-PFE solution for this policy). This is of course 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 the dates of either t or T. If we adopt the alternative numbering scheme $\tau \equiv T - t$ (i.e., we number periods according to the number remaining until the shift to the long-run policy), then the solution for $e_\tau(n)$ for any $\tau \geq 1$ will be independent of T. Moreover, in terms of this notation, the belief-revision dynamics (2.18) can be written in the form

$$\dot{e}_\tau(n) = -e_\tau(n) + \sum_{j=1}^{\tau-1} [\psi_j e_{\tau-j}(n) + \varphi_{j2} \bar{\imath}_{SR}] + \sum_{j=\tau}^{\infty} [\psi_j e_{LR}(n) + \varphi_{j2} \bar{\imath}_{LR}]$$

for each $\tau \geq 1$, where φ_{j2} is the second column of the matrix φ_j. These dynamics can equivalently be written in the form

$$\begin{aligned}
\dot{e}_\tau(n) = & -e_\tau(n) + (I - \Lambda) \sum_{j=1}^{\tau-1} \Lambda^{j-1} [M e_{\tau-j}(n) + m_2 \bar{\imath}_{SR}] \\
& + \Lambda^{\tau-1} [M e_{LR}(n) + m_2 \bar{\imath}_{LR}],
\end{aligned} \tag{3.6}$$

and integrated forward from the initial conditions $e_\tau(0) = 0$ for all $\tau \geq 1$, using solution (3.3) for $e_{LR}(n)$.

We observe that for $\tau = 1$, the linear differential equation (3.6) can be solved uniquely for the function $e_1(n)$, given that $e_{LR}(n)$ is already known. Then the equation for $\tau = 2$ can be solved uniquely for the function $e_2(n)$, given that $e_1(n)$ and $e_{LR}(n)$ are already known; and proceeding recursively in this way, one can solve uniquely for the $\{e_\tau(n)\}$ for all values of τ up to any

28

given bound T (corresponding to the initial period $t = 0$). In this way, we obtain a unique solution for $e_t(n)$ for all $t \geq 0$.

Note further that considering how $e_t(n)$ changes (for any fixed t) as T is increased is equivalent to considering how the solution to the system of differential equations (3.6) changes for progressively larger values of τ. In particular, the behavior of $e_t(n)$ as T is made unboundedly large can be determined by calculating the behavior of the solution to the system (3.6) 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 (2.7), where the coefficients (ϕ_π, ϕ_x) are constant over time and such that the left-hand side of (2.13) 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

$$e_{SR}(n) = [I - \exp[n(M - I)]] \, \bar{e}_{SR}^{PF}, \tag{3.7}$$

where

$$\bar{e}_{SR}^{PF} \equiv [I - M]^{-1} m_2 \, \bar{\imath}_{SR}. \tag{3.8}$$

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 TE relations (2.9) and the reaction function (2.7).

The proof is given in 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. Comparison of expressions (3.7)–(3.8) with (3.3)–(3.4) also shows that the predicted outcomes in the case of any long enough horizon T for maintenance of the "temporary" policy are close to the predicted outcomes (under a reflective equilibrium with the same degree of reflection n) in the case that the policy is expected to be permanent. 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. If the coefficients (ϕ_π, ϕ_y) satisfy the Taylor Principle (2.13), then (3.5) implies that as n is made large,

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

Moreover, the beliefs $\bar{e}_{SR}^{PF}$ defined in (3.8) are simply the steady-state PFE beliefs (or FS-PFE beliefs) in the case of a permanent commitment to the reaction function (2.7) with $\bar{\imath}_t = \bar{\imath}_{SR}$. Thus we obtain the following result.

Proposition 3 *Suppose that in addition to the hypotheses of Proposition 2, the coefficients (ϕ_π, ϕ_y) satisfy the Taylor Principle (2.13). Then the limits*

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

and

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

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 (2.7) with $\bar{\imath}_t = \bar{\imath}_{SR}$.

Proposition 3 identifies a case in which the thought experiment of considering the PFE consistent with 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 (if one selects the FS-PFE solution, as is conventional in the NK literature), but this answer provides a good approximation
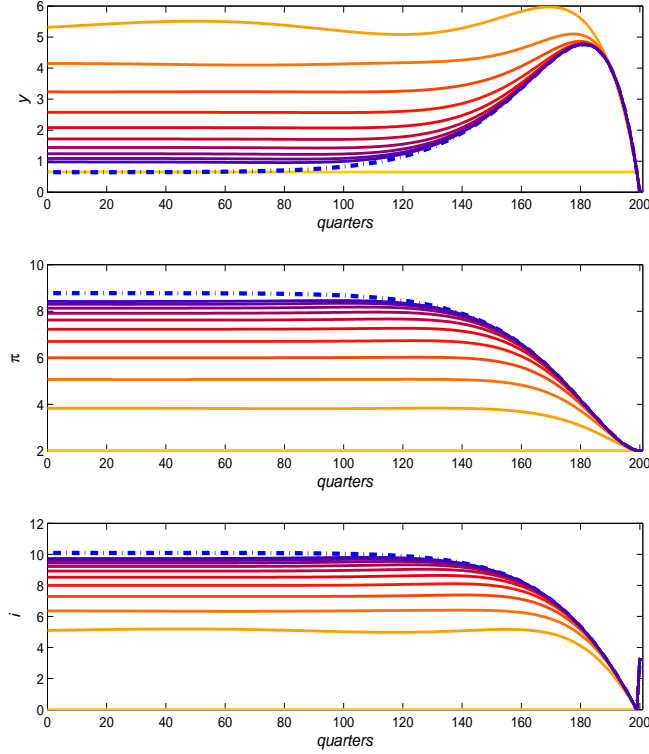
Figure 2: Reflective equilibrium outcomes for $n = 0$ through 20 (progressively darker lines) compared with the FS-PFE solution (dash-dotted line), when the Taylor-rule intercept is reduced for 200 quarters.

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.

Figure 2 provides a numerical illustration of these results. The policy experiment is the same as in Figure 1, as are the assumed numerical parameter values, except that in Figure 2 the commitment to the intercept $\bar{\imath} < 0$ is expected to last for 50 years. Again the reflective equilibrium paths are shown for progressively higher values of n.[38] The figure shows not only the convergence of the reflective equilibrium outcomes for all three variables as T is made large, for each of the possible values of n, but also the convergence of reflective

---

[38] Again, the movement from lighter to darker lines corresponds to increasing n. The lines shown in the figure correspond to the values $n = 0, 2, 4$, and so on up to $n = 20$.

equilibrium to the FS-PFE predictions, for each of the possible values of τ (and hence for each possible value of T).

Figure 2 shows that the FS-PFE provides a good approximation to the reflective equilibrium outcome if the degree of reflection is on the order of $n = 20$; but this would involve a degree of reflection that seems fairly unrealistic, if it is taken to represent a purely prospective calculation on the part of people who have learned about an announced policy change, but not yet had occasion to observe what actually happens under the new regime.

This illustrates a general point: in judging the practical relevance of the PFE prediction, it matters not only whether reflective equilibrium should converge to the PFE as n is made large enough, but also how *quickly* such convergence should occur. The speed of convergence is not too great an issue in the case of a commitment to a new policy to be maintained for only a few quarters, when the new policy is a Taylor rule, and when the temporary policy is to be followed by a reversion to a policy regime that the public already understands well on the basis of past experience (the experiment considered in Figure 1). It is a bigger issue, however, in the case of a commitment to a new permanent regime that differs non-trivially from past policy, even when both the old and the new policies conform to the Taylor principle. And, as we show in the next section, it is a still larger issue in the case of a temporary regime under which the Taylor principle is not expected to be satisfied, as in the case of a commitment to a fixed interest rate for a significant period of time.

We can now address a question posed in the title of our paper: what should happen if people come to expect that a "loose" monetary policy will be maintained for several more years? Our results show that in our model, an expectation that the reaction-function intercept will be kept lower than usual for the next several years should lead to *higher* inflation and output, regardless of the degree of reflection, and regardless of the length of time for which the looser policy is expected to be maintained. If the loosening of policy is expected to be sufficiently transitory (though it may last for some years), as in Figure 1, then the policy change will be associated with a temporarily lower nominal interest rate, regardless of the degree of reflection. But if the shift in the policy rule is expected to last for a sufficiently long (though possibly

finite) period of time, the higher inflation rate and output will be associated with a *higher* nominal interest rate, despite the reduction in the intercept of the central bank's reaction function, except in the case of a degree of reflection that is very low.[39]

4 Consequences of a Temporarily Fixed Nominal Interest Rate

We now consider the case in which it comes to be understood (either as a result of a shock, or a policy announcement) 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.[40] 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 for some time; 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, regardless of the exact outcomes for output and inflation, 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. The "date-based forward guidance" provided by several central banks in the aftermath of the global financial crisis arguably involved commitments of this kind; and while no explicit promises were made about how policy would be conducted *beyond* the horizons in question, one might suppose that people would expect the central bank to revert to its usual approach to policy once there ceased to be any explicit commitment to behave otherwise. We are interested in the extent to which such a temporary change in policy should have effects

---

[39]For a discussion about a permanent shift in policy and the satisfaction of the Fisher Equation, refer to García-Schmidt and Woodford (2015, sec. 3.4).

[40]The latter policy is assumed to be a rule of the form (2.7), in which the response coefficients satisfy the Taylor Principle (2.13), and the intercept is consistent with the inflation target π^*.

similar or different from the effects of a temporary shift in the intercept of the monetary policy reaction function, analyzed above.

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 (perhaps by the expected persistence of an exogenous real disturbance). 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 it might actually be expansionary to do so. A second question is the effect of changing T, 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 (which may be infeasible due to the ZLB)?

4.1   Convergence to Perfect-Foresight Equilibrium

We first consider whether reflective equilibrium converges to a PFE again in this case, as n grows, and if so to which of the possible PFE paths. The question of equilibrium selection is of particular interest in this policy experiment, since here, unlike the case considered in section 3, different criterion to choose among the PFE imply a different equilibrium selection.[41]

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. If we suppose that $g_t = 0$ for all $t \geq T$, this means that the reflective equilibrium outcomes for all $t \geq T$ will converge to the steady state consistent with the inflation target π^*. Note that this simple result already tells us that if the reflective equilibrium converges to *any* PFE, it can only converge to the FS-PFE.

---

[41]For example, the "backward stability" selection criterion proposed by Cochrane (2016) selects a different equilibrium than the PFE-FS.

The analysis of convergence prior to date T requires an extension of our previous result, because now we assume that the response coefficients (ϕ_π, ϕ_y) differ before and after date T. Nonetheless, as shown in 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 (2.7) for all $t \geq T$, where the coefficients (ϕ_π, ϕ_x) of the latter function satisfy (2.13), and the intercept is consistent with the inflation target π^*. Then in the case of any initial conjecture $\{e_t(0)\}$ regarding average expectations that converges exponentially, the belief revision dynamics (2.18) 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. The model parameters are as in the previous numerical examples, and for simplicity we again show the effects of a pure shift in monetary policy, assuming $g_t = 0$ for all t and an initial conjecture under which $e_t(0) = 0$ for all t. As in Figure 1, it is again assumed that monetary policy is expected 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 1: the two shifts in monetary policy have been chosen to lower the nominal interest rate to the same extent (i.e., 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 1,
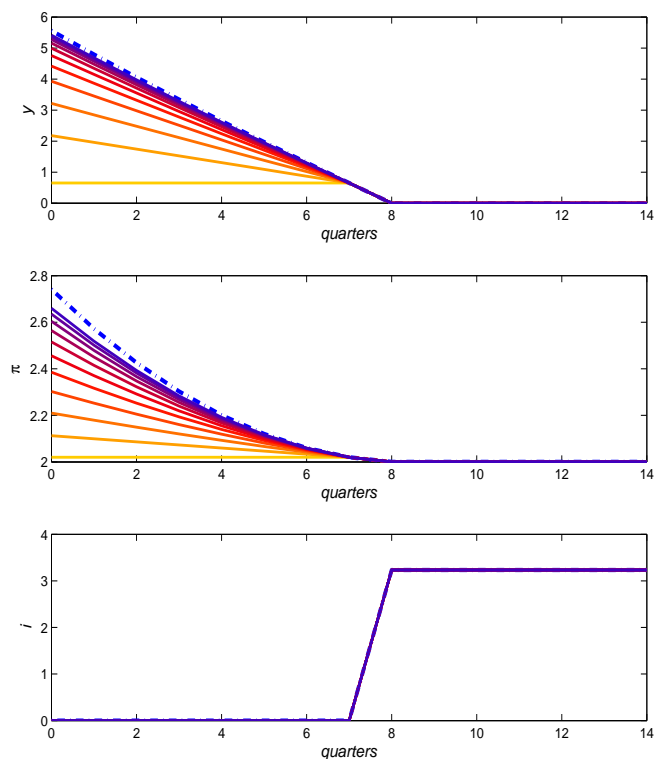
Figure 3: Reflective equilibrium outcomes for $n = 0$ through 4 compared with the FS-PFE solution, when the nominal interest rate is fixed for 8 quarters.

but not exactly the same: the output and inflation increases are somewhat larger when the interest rate is expected to remain fixed, because now there is no expectation of endogenous interest-rate increases in subsequent periods in response to the increases in output and inflation.

The effects are larger the greater the degree of reflection, and strongest under the assumption of perfect foresight. (They are also larger the longer the time for which the interest rate is expected to remain fixed.) Also, the difference between the PFE predictions and those obtained from a given finite degree of reflection is greater than that obtained in the case of a temporary shift in the Taylor-rule intercept.

In Figure 3, as in Figure 1, an average degree of reflection of $n = 4$ results

36

in TE outcomes that are similar to the PFE predictions. But the reflective equilibrium outcomes when $n = 2$ are not as close to the PFE outcomes as they are in Figure 1, especially in the first quarters. In quarter zero, the output response when $n = 2$ is 14 percent smaller than the PFE 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 PFE predictions. Moreover, these discrepancies rapidly become much larger if the interest rate is expected to be fixed for an even longer period of time.

4.2 Very Long Periods with a Fixed Nominal Interest Rate

Much recent criticism of the implications of standard New Keynesian models regarding the effects of "forward guidance" has focused on the implications of such models if one assumes that the nominal interest rate would be fixed for several years.[42] It should be noted that no central banks have actually experimented with date-based forward guidance that referred to dates more than about two years in the future; and while the period in which the U.S. federal funds rate target remained at its lower bound lasted for seven years, there was little reason for anyone to expect it to remain at this level for so long when the lower bound was reached at the end of 2008. Nonetheless, as discussed in section 1, thought experiments involving long-lasting periods at the ZLB remain useful for clarifying the theoretical coherence of proposed solution concepts.

If one assumes a date T many years in the future, the FS-PFE predicted effects on both output and inflation rapidly become extremely large. However, the effects predicted by reflective equilibrium with some modest (though positive) degree of reflection n do not grow in the same way, so that the PFE prediction rapidly becomes a worse and worse approximation to what one should expect in a reflective equilibrium, if the horizon T is very long. Figure 4 illustrates this, in the case of the same model calibration as used in previous

---

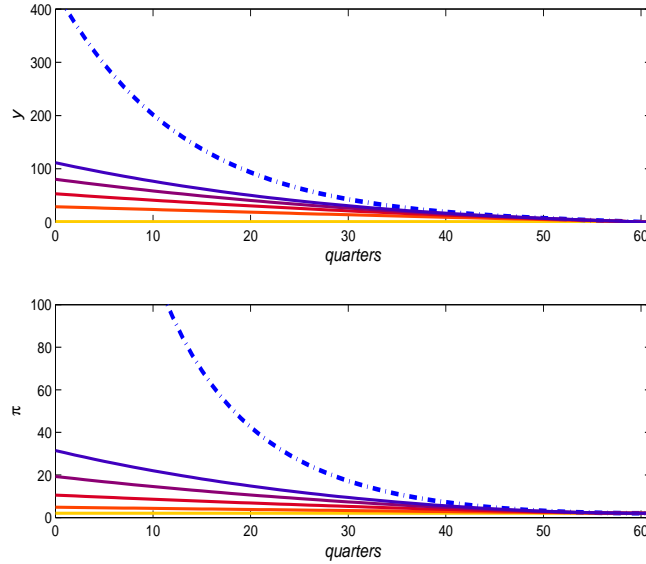[42]See, for example, Chung (2015), Del Negro et al. (2015), McKay et al. (2016), and Cochrane (2016).

Figure 4: Reflective equilibrium outcomes for $n = 0$ through 4 compared with the FS-PFE solution, when the nominal interest rate is fixed for 15 years.

figures, by considering a (certainly unrealistic) situation in which the nominal interest rate is expected to be fixed for 15 years.[43]

According to the log-linearized model, an expectation of remaining at the ZLB for such a long time would, under the FS-PFE analysis, imply extremely large effects in the initial quarter: log output higher than its steady-state level by 4.36 (output 78 times its steady-state level), and an inflation rate of 442 percent.[44]

In the case of the reflective equilibrium, even if we assume $n = 4$ (a rather high average degree of reflection), the predicted increase in log output in quarter zero is instead only 1.11 (output 3 times its steady-state level), while the

---

[43]The third panel of the figure is omitted, since the expected path of the nominal interest rate is independent of the degree of reflection, as in Figure 3. The figure shows the degrees of reflection $n = 0, 1, 2, 3, 4$.

[44]Such extreme predictions make it foolish to believe the assumptions made in this calculation (even given the assumption about policy): log-linearization of the model will not yield even a roughly accurate result, nor do even the assumptions of the exact NK model — $1 - \alpha$ of firms not reconsidering their prices. We mention them only to point out that even granting the validity of our log-linearized model for purposes of such an exercise, the FS-PFE predictions are not at all a close approximation to the reflective equilibrium predictions.

inflation rate is predicted to increase only to 31.5 percent per annum. If we assume a more modest degree of reflection, $n = 2$, the predicted increase in log output is only 0.53 (output 1.7 times its steady-state level), and inflation is predicted to increase only to 10.6 percent. This is still quite a large increase in output (large enough to make one doubt the realism of using the model for such an analysis), but these results are not close to the shocking predictions of the FS-PFE analysis.

The FS-PFE predictions of the log-linearized model become even more extreme if a longer period at the ZLB is contemplated: both the predicted effects on output and inflation grow without bound (and quite rapidly) as T is increased. This is because the matrix B mentioned in section 2.2 has an eigenvalue greater than 1 and so output and inflation grow exponentially as T is increased.[45]

One of the objections to the FS-PFE is based on this ground, noting that it is implausible to suppose that changes in the specification of policy only very far in the future (say, a commitment to maintain the low interest rate for 1001 quarters instead of for only 1000 quarters) should have any significant effect on current economic outcomes.[46] But this unpalatable feature of the FS-PFE is not a property of our concept of reflective equilibrium, assuming a fixed degree of reflection n as the length of the policy commitment is increased. 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 (3.7), where $\bar{e}_{SR}^{PF}$ is

---

[45]Note further that the elements of x_t are the *logarithm* of output and the *continuously compounded* rate of inflation; these quantities must both be exponentiated to obtain the level of output and the factor by which prices increase relative to the pevious year's prices.

[46]See Cochrane (2016).

again defined in (3.8). 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 TE relations (2.9) and the reaction function (2.7).

The proof is given in the Appendix. The result is similar to the one stated in Proposition 2.[47] There is one important difference, however: in the present case, the stationary expectations $\bar{e}_{SR}^{PF}$ no longer correspond to a unique FS-PFE associated with permanent maintenance of the interest rate $i_t = \bar{\imath}_{SR}$.

Thus if we consider the reflective equilibrium associated with any given finite degree of reflection n, we find that equilibrium outcomes are essentially the same for any long enough horizon T. Thus there is no material difference, between commitment to a fixed interest rate for a long but finite time and a commitment to fix the interest rate permanently.[48]

An associate question is the effects of maintaining a low interest rate for longer. As is further discussed and shown in García-Schmidt and Woodford (2015, sec.4.4), for any degree of reflection, the commitment to keep the nominal interest rate at a low level for longer is both expansionary and inflationary as is discussed in the following proposition:

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 $t \geq T$.[49] Then for any fixed $\bar{\imath}_{SR}$ and fixed level of reflection $n > 0$, increasing the length of the commitment from*

---

[47]Note that Proposition 2, as stated earlier, did not require that the reaction function coefficients satisfy (2.13); it would apply, in particular, to the case $\phi_\pi = \phi_y = 0$, corresponding to fixed interest rates before and after date T. The only difference here is that Proposition 5 establishes a similar result even when the response coefficients prior to date T differ from those from date T onward.

[48]Refer to García-Schmidt and Woodford (2015, sec. 4.2) for a discussion of similarities between the reflective equilibrium and other proposed equilibrium selection mechanisms.

[49]In fact, it should be evident from the proof given in 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.

*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 the Appendix. Even when the degree of reflection is
very low, the outcomes qualitatively resemble those predicted by the FS-PFE
analysis, even if the quantitative magnitude of the effects is quite different.

It is important to note however that with the reflective equilibrium when
there is a negative shock to the economy, even maintaining a low fixed inter-
est forever may not prevent the economy of entering a recession. While the
FS-PFE analysis implies that the effects of *any* contractionary shock, can nec-
essarily be completely counter-acted by a sufficiently long-lasting commitment
to a low interest rate — and in fact, that a sufficiently long-lasting commit-
ment 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.

4.3 The Paradox Explained

We can now explain the error in the reasoning sketched in the introduction. It
is true that under the assumption of a *permanent* interest-rate peg, the only
forward-stable PFE are ones that converge asymptotically to an inflation rate
determined by the Fisher equation and the interest-rate target (and thus, lower
by one percentage point for every one percent reduction in the interest rate).
But for most possible initial conjectures (as starting points for the process
of belief revision proposed 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 PFE;
instead, one generally finds that the belief-revision dynamics *fail to converge*
to *any* PFE as n increases, in the case of a permanent interest-rate peg.

This failure of convergence can be illustrated using results already pre-

sented above. In the case of a policy under which $i_t = \bar{\imath}_{LR}$ forever, if we 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 (3.2) for all t, where M in this equation is now the matrix corresponding to response coefficients $\phi_\pi = \phi_y = 0$, and we now have $e_t(n) = e_{LR}(n)$ for all t.[50] The solution for general n is again given by (3.3), where $\bar{e}_{LR}^{PF}$ is again defined by (3.4). However, whereas in the Taylor-rule case considered in section 3, this solution implied that $e_{LR}(n) \to \bar{e}_{LR}^{PF}$ 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 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 $\bar{e}_{LR}^{PF}$, rather than converging to it. Nor does $e_{LR}(n)$ approach any PFE: 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

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

or

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

is well-defined in the case of a temporary interest-rate peg.[51] It is true (for any finite length of peg) that a high enough degree of reflection leads to an outcome indistinguishable from a forward-stable 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

---

[50]The case considered now is of the same kind as was considered in deriving (3.2), except that we now set $T = 0$, and assume $\phi_\pi = \phi_y = 0$.

[51]Note that $\bar{e}_{SR}^{PF}$, the common limit given in Proposition 3, 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} converge to it as T is made large. Failure of the "Taylor Principle" invalidates *both* of those convergence results, relied upon in Proposition 3.

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 PFE 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.[52] 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 (2012); but it is not an inherent problem with announcing a change in future policy intentions, only with a particular way of explaining what has changed.

The idea that a commitment to keep nominal interest rates low for a longer time should be deflationary, even when understood to represent a pure change in monetary policy — simply because the only rational-expectations equilibria in which nominal interest rates remain forever low involve deflation — is instead mistaken, in our view. 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 interest rate, it does not follow that they must expect a deflationary equilibrium; this does not follow, even if we suppose that they reason about the economy's likely future path using a correct model of how inflation and aggregate output are determined.

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

---

[52]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).

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 the outcome associated with 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 PFE, even approximately.

This is why it is important to explicitly model the process of belief revision as a result of further reflection, rather than simply assuming that the PFE must yield a correct prediction. Some macroeconomists may find the proposed alternative solution concept (reflective equilibrium for some finite degree of reflection n) unappealing, on the ground that it yields a less definite prediction than the assumption of perfect foresight (or rational expectations) equilibrium. But while it is true that our conclusions about the effects of a given policy commitment depend both on the exact choice of an initial conjecture and on exactly how far one supposes that people should continue the belief-revision process, this does not mean that we are unable to draw any conclusions of relevance to policy deliberations. Our conclusions as to the *signs* of the effects just mentioned are independent of those details of the specification of the reflective equilibrium. 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 (for differing assumptions about the initial conjecture and the degree of reflection) 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 of policy design to choose a policy with the property that the belief-revision dynamics should converge reliably, leading to less uncertainty about the outcome that should be expected under the policy.

In the case of a central bank that finds itself seeking additional demand stimulus when it has already cut its short-term nominal interest rate instrument to its effective lower bound, a commitment to maintain the instrument at the lower bound for a long time *that can be announced in advance,* 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, according to the analysis in this paper; 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 to the ultra-low interest rate that does not run simultaneously the risk of being too short to be effective, if the degree of reflection n is too low, and the opposite risk of being 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), or the nominal-GDP target path proposed by Woodford (2012).[53] 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, under such a commitment; 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, this does not mean that the approach yields no conclusions that are useful for policy design. Instead, insisting on the use of perfect foresight equilibrium analysis simply because it yields a more precise prediction may lead to large errors. One is reminded of the dictum of the British logician Carveth Read:[54] "It is better to be vaguely right than exactly wrong."

---

[53]This alternative to date-based forward guidance would also have the advantage of being less likely to be misunderstood as revealing negative central-bank information about fundamentals, as discussed by Woodford (2012).

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

References

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.

Chung, Hess (2015). "The Effects of Forward Guidance in Three Macro Models". FEDS Notes, Federal Reserve Board, February 26, 2015.

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

— (2014). "Monetary Policy with Interest on Reserves". *Journal of Economic Dynamics & Control* 49, pp. 74–108.

— (2015). "Doctrines Overturned". The Grumpy Economist, February 28, [URL: http://johnhcochrane.blogspot.com/2015/02/doctrines-overturned.html].

— (2016). "The New-Keynesian Liquidity Trap". unpublished, revised June.

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 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.

Friedman, Milton (1968). "The Role of Monetary Policy". *American Economic Review* 58 (1), pp. 1–17.

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.

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

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

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

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.

— (2012). "Methods of Policy Accommodation at the Interest-Rate Lower Bound". *Proceedings - Economic Policy Symposium - Jackson Hole*, pp. 185–288.

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

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

APPENDIX

A Matrices of Coefficients and their Properties

A.1 Temporary Equilibrium Solution

The system of three equations given in the text can be solved to obtain

$$x_t = Ce_t + c\omega_t \qquad (A.1)$$

where we define the vectors

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

and the matrices

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

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 (2.7).

This solution also allows us to solve for the summary variables a_t that decisionmakers need to forecast, resulting in

$$a_t = Me_t + m\omega_t$$

where we define

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

A.2 Perfect Foresight Equilibrium Dynamics

It follows from the discussion in the text (citing Woodford 2003, chap. 4) that the PFE dynamics can be written in the form

$$x_t = B\, x_{t+1} + b\,(\rho_t - \bar{\imath}_t) \qquad (A.2)$$

49

where we define

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

Alternatively, we can characterize PFE dynamics by the requirement that e_t must equal e_t^* for all t. From (2.17) it follows that a sequence of vectors of expectations $\{e_t\}$ constitute PFE expectations if and only if

$$
\begin{aligned}
e_t &= e_t^* = \sum_{j=1}^{\infty} \psi_j e_{t+j} + \sum_{j=1}^{\infty} \varphi_j \omega_{t+j} \\
&= \psi_1 e_{t+1} + \varphi_1 \omega_{t+1} + \Lambda\, e_{t+1} \\
&= (I - \Lambda) M\, e_{t+1} + (I - \Lambda) m\, \omega_{t+1} + \Lambda\, e_{t+1} \\
&= [(I - \Lambda) M + \Lambda]\, e_{t+1} + (I - \Lambda) m\, \omega_{t+1} \qquad (A.3)
\end{aligned}
$$

for all $t \geq 0$.

The dynamics implied by (A.3) are in fact equivalent to those implied by (A.2). Using (A.1) together with (A.3) implies that the PFE dynamics of output and inflation must satisfy

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

But this relation is in fact equivalent to (A.2), given that our definitions above imply that

$$C\,[(I - \Lambda) M + \Lambda]\, C^{-1} = B, \qquad (A.4)$$

$$C\,(I - \Lambda) m = Bc + b \cdot [-\beta\sigma^{-1} \ \ 0],$$

$$c = b \cdot [\sigma^{-1} \ \ -1].$$

A.3   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

$$\phi_\pi + \frac{(1-\beta)}{\kappa}\phi_y - 1 \;\neq\; 0. \tag{A.5}$$

(In this case the steady-state vector of expectations (3.4) 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 $\text{Tr}[A] < 0$.[55] From the result above, the first of these conditions is satisfied if the left-hand side of (A.5) is positive, which is to say, if the Taylor Principle (2.13) 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

$$\phi_y + \kappa\phi_\pi - \kappa \;=\; \kappa\left(\phi_\pi + \frac{\phi_y}{\kappa} - 1 \right) \;>\; \kappa\left(\phi_\pi + \frac{\phi_y(1-\beta)}{\kappa} - 1 \right) \;>\; 0. \tag{A.6}$$

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 (A.5) is negative, $\text{Det}[M - I] < 0$, and as a consequence the matrix must have two real eigenvalues of opposite sign.[56] Thus one eigenvalue is positive in this case, as asserted in the text. Note that this is the case that obtains if $\phi_\pi = \phi_y = 0$.

---

[55]See, for example, Hirsch and Smale (1974, p. 96).
[56]Again see Hirsch and Smale (1974, p. 96).

A.4 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 (2.13) 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

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

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 $\mathrm{Det}[A(\lambda)M-I] > 0$ for all $|\lambda| \leq 1$, if and only if the Taylor Principle is satisfied.

The trace of the matrix is given by

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

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 (A.6). 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.

A.5   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

$$\Delta\,\mathrm{Det}\,B \;=\; \beta, \tag{A.7}$$

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

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

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

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

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

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

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 (2.13) 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

$$(\mu_1 - 1)(\mu_2 - 1) < 0. \tag{A.8}$$

This condition is inconsistent with the eigenvalues being a pair of complex conjugates, so in this case there must be two real eigenvalues. Condition (A.8) further implies that one must be greater than 1, while the other is less than 1. Condition (A.7) 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 (A.5) 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

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

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 (A.5) 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 (A.2) is then given by $x_t = \bar{x}$ for all t, where

$$\bar{x} \equiv (I - B)^{-1}b\left[(1 - \beta)\sigma^{-1}\bar{g} - \bar{\imath}\right]. \tag{A.9}$$

Note that condition (A.5) 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 (A.3) 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 (A.9) because of the identities linking the M and B matrices noted above.

A.6 Convergence of the PFE Dynamics

As noted in the text in section 2.2, in the case that $\phi_\pi = \phi_y = 0$, there exists a continuum of PFE solutions that remain bounded for all t, described by equations (2.15) for alternative values of the coefficient χ. Here we show that if after some finite date T, both $\bar{\imath}_t$ and ρ_t take constant values, then each of this continuum of solutions has the property that

$$\lim_{t \to \infty} \pi_t = \pi_{LR}, \qquad \lim_{y \to \infty} y_t = y_{LR},$$

where the limiting values are independent of χ and are given by (2.16). Moreover, the limiting values to which the PFE dynamics converge correspond to the PFE steady state (A.9).

If $\bar{\imath}_t = \bar{\imath}_{LR}$ and $\rho_t = \rho_{LR}$ for all $t \geq T$, then for any $t \geq T$, (2.15) takes the form

$$
\begin{aligned}
x_t &= v_1(e_1'b) \frac{\rho_{LR} - \bar{\imath}_{LR}}{1 - \mu_1} - v_2(e_2'b) \left\{ \sum_{j=1}^{t-T} \mu_2^{-j} \cdot (\rho_{LR} - \bar{\imath}_{LR}) + \sum_{j=t+1-T}^{t} \mu_2^{-j} (\rho_{t-j} - \bar{\imath}_{t-j}) \right\} \\
&\quad + \chi v_2 \mu_2^{-t} \\
&= \left[\frac{v_1(e_1'b)}{1 - \mu_1} - \frac{v_2(e_2'b)}{\mu_2 - 1} (1 - \mu_2^{T-t}) \right] \cdot (\rho_{LR} - \bar{\imath}_{LR}) + C v_2 \mu_2^{-t} + \chi v_2 \mu_2^{-t}
\end{aligned}
$$

where

$$C \equiv \sum_{s=0}^{T-1} \mu_2^s (\rho_s - \bar{\imath}_s)$$

has a value independent of t. Given that $0 < \mu_2^{-1} < 1$, we see immediately from this that x_t converges to

$$x_{LR} \equiv \left[\frac{v_1(e_1'b)}{1 - \mu_1} + \frac{v_2(e_2'b)}{\mu_2 - 1} \right] \cdot (\rho_{LR} - \bar{\imath}_{LR})$$

as $t \to \infty$. This limiting vector is independent of the value of χ.

Finally, we note that

$$
\begin{aligned}
(I - B)\, x_{LR} \;&=\; (I - B) \left[\frac{v_1(e_1'b)}{1 - \mu_1} - \frac{v_2(e_2'b)}{\mu_2 - 1} \right] \cdot (\rho_{LR} - \bar{\imath}_{LR}) \\
&=\; \left[\frac{(1 - \mu_1)v_1(e_1'b)}{1 - \mu_1} - \frac{(1 - \mu_2)v_2(e_2'b)}{\mu_2 - 1} \right] \cdot (\rho_{LR} - \bar{\imath}_{LR}) \\
&=\; [v_1(e_1'b) + v_2(e_2'b)] \cdot (\rho_{LR} - \bar{\imath}_{LR}) \\
&=\; b \cdot (\rho_{LR} - \bar{\imath}_{LR}),
\end{aligned}
$$

so that x_{LR} is just the vector of steady-state values defined in (A.9). Our definitions of B and b above further imply that when $\phi_\pi = \phi_y = 0$,

$$
(I - B)^{-1}\, b \;=\; \left[\begin{array}{c} -\frac{1-\beta}{\kappa} \\ -1 \end{array} \right],
$$

so that (A.9) implies the values given in (2.16).

B   Proofs of Propositions

B.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 (2.18) 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 (2.17) 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 $t < \bar{T}$,

$$e_t^*(n) \;=\; \sum_{j=1}^{\bar{T}-t-1} [\psi_j e_{t+j}(n) + \varphi_j \omega_{t+j}] \;+\; \Lambda^{\bar{T}-t-1}[Me_\infty(n) + m\omega_\infty]$$

$$+\; \sum_{k=1}^{K} \lambda_k^{-1}\Lambda^{\bar{T}-t-1}A(\lambda_k)\,[Ma_{e,k}(n) + ma_{\omega,k}].$$

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 (2.18) can be written in the more compact form

$$\dot{\mathbf{e}}(n) \;=\; V\,\mathbf{e}(n) \;+\; W\,\boldsymbol{\omega}, \tag{B.10}$$

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

$$V = \begin{bmatrix} V_{11} & V_{12} \\ 0 & V_{22} \end{bmatrix},$$ (B.11)

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

$$V_{11} = \begin{bmatrix} -I & v_{12} & \cdots & v_{1,\bar{T}-1} & v_{1,\bar{T}} \\ 0 & -I & \cdots & v_{2,\bar{T}-1} & v_{2,\bar{T}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -I & v_{\bar{T}-1,\bar{T}} \\ 0 & 0 & \cdots & 0 & -I \end{bmatrix},$$ (B.12)

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

$$V_{22} = \begin{bmatrix} A(\lambda_1)M - I & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & A(\lambda_K)M - I & 0 \\ 0 & \cdots & 0 & M - I \end{bmatrix}.$$ (B.13)

These results allow us to determine the eigenvalues of V. The block-triangular form (B.11) 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 (B.12) 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 (B.13) 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 A.3, it follows from the hypothesis that the reaction function coefficients satisfy (2.13) 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 (B.10), defined by:

$$\mathbf{e}^{PF} \equiv -V^{-1}W\boldsymbol{\omega}.$$ (B.14)

It also implies that the dynamics (B.10) 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).[57]

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 (B.10).

Finally, the paths of output and inflation in any reflective equilibrium are given by (A.1), given the solution for $\{e_t(n)\}$. Using (2.7), 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. 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.

B.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 (3.3). 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

$$
\dot{e}_\tau(n) = -e_\tau(n) + (I - \Lambda) \sum_{j=1}^{\tau-1} \Lambda^{j-1} [Me_{\tau-j}(n) + m_2 \bar{\imath}_{SR}]
$$
$$
+ \Lambda^{\tau-1} [Me_{LR}(n) + m_2 \bar{\imath}_{LR}] \tag{B.15}
$$

---

[57]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 (B.10) apply at all times; this depends on assuming that the reaction function (2.7) can be implemented at all times, which requires that the zero lower bound never bind. Thus we only establish convergence for all those initial conjectures such that the dynamics implied by (2.18) never cause the zero lower bound to bind. There is however a large set of initial conditions for which this is true, given that the unconstrained dynamics are asymptotically convergent.

derived in the text, together with the initial conditions $e_\tau(0) = 0$ for all $\tau \geq 1$. (Equation (B.15) repeats equation (3.6) 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

$$X_n(z) \equiv \sum_{\tau=1}^{\infty} e_\tau(n)\, z^{1-\tau}. \tag{B.16}$$

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 (B.16) with respect to n, and substituting (B.15) for $\dot{e}_\tau(n)$ in the resulting equation, we obtain an evolution equation for the z-transform:

$$
\begin{aligned}
\dot{X}_n(z) &= -\sum_{\tau=1}^{\infty} e_\tau(n)z^{1-\tau} + (I - \Lambda)\sum_{j=0}^{\infty}\Lambda^j z^{-j}\left[M\sum_{\tau=1}^{\infty} e_\tau(n)z^{-\tau} + m_2\bar{\imath}_{SR}\sum_{\tau=1}^{\infty} z^{-\tau}\right] \\
&\quad + \sum_{j=0}^{\infty}\Lambda^j z^{-j}\left[M\, e_{LR}(n) + m_2\bar{\imath}_{LR}\right] \\
&= -X_n(z) + (I - \Lambda)(I - \Lambda z^{-1})^{-1}\left[z^{-1}M\, X_n(z) + (z-1)^{-1}m_2\bar{\imath}_{SR}\right] \\
&\quad + (I - \Lambda z^{-1})^{-1}\left[M e_{LR}(n) + m_2\bar{\imath}_{LR}\right], \tag{B.17}
\end{aligned}
$$

which holds for any $n > 0$ and any z in the region of convergence. We note that the right-hand side of (B.17) 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 (B.17) 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 (B.17) is independent for each value of z. (This is the advantage of z-transformation of the original system of equations (B.15).) 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

$$
\begin{aligned}
X_n(z) \;=\; & (1 - z^{-1})^{-1} \left[I - \exp(n(M - I)) \right] (I - M)^{-1} \cdot m_2 \bar{\imath}_{LR} \\
& + (z - 1)^{-1} \left[I - \exp(-n\Phi(z)) \right] \Phi(z)^{-1} (I - \Lambda)(I - \Lambda z^{-1})^{-1} \\
& \qquad\qquad\qquad\qquad \cdot m_2(\bar{\imath}_{SR} - \bar{\imath}_{LR}) \qquad\qquad\qquad\text{(B.18)}
\end{aligned}
$$

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 (B.18) 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[58] 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 (B.18), we observe that the limit is well-defined, and equal to

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

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.

B.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 (2.13) is satisfied, then as shown in section A.3 above, both eigenvalues of $M - I$ have negative real part. From this (3.5) follows; substituting of this

---

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

into (3.7) yields

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

where $\bar{e}_{SR}^{PF}$ is defined in (3.8). This establishes the first double limit in the statement of the proposition.

The result that

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

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

As discussed in section A.2 above, the FS-PFE dynamics $\{e_t^{PF}\}$ satisfy equation (A.3) 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, $e_t^{PF} = \bar{e}_{LR}^{PF}$ for all $t \geq T - 1$, where

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

is the same as the vector defined in (3.4).

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

$$e_\tau = [(I - \Lambda)M + \Lambda] e_{\tau-1} + (I - \Lambda) m_2 \bar{\imath}_{SR} \qquad (B.19)$$

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

$$(I - \Lambda)M + \Lambda = C^{-1}BC, \qquad (B.20)$$

which must be the same as the eigenvalues of B. (Note that (B.20) follows from (A.4).)

Under the hypothesis that the response coefficients satisfy the Taylor Principle (2.13), both eigenvalues of B are inside the unit circle. It then follows that the dynamics (B.19) converge as $\tau \to \infty$ to the steady-state vector of expectations $\bar{e}_{SR}^{PF}$ defined in (3.8). We thus conclude that

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

for any t. This establishes the second double limit.

B.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 (B.10), 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 (B.11)–(B.13). In equation (B.13), the matrix M is defined using the coefficients (ϕ_π, ϕ_y) that apply after date T, and thus that satisfy the Taylor Principle (2.13), 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 (B.10) 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, 2016) 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 (B.10), and the matrix V will again have the form (B.11)–(B.13). But in this case, the matrix M in (B.13) 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 (B.10)

will be explosive in the case of almost all initial conjectures (even restricting our attention to conjectures within the specified finite-dimensional family).

B.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 $t \geq T$.[59] 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 satisfy (A.5). 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 (3.7)–(3.8). This establishes the proposition.

B.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

---

[59]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 $t \geq T$.

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 $i = 1, 2$).[60]

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 (A.1), 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 some 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 $a_t^2 - a_t^1 >> 0$, 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 $t = T' - 1$.

---

[60]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 $\bar{\imath}_{SR}$.

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 $T \le t < T' - 1$, we continue to have $e_t^{*1}(n) = 0$ for all n, for the same reason as in the case of periods $t \ge T'$. 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 $T \le t < T' - 1$.

Moreover, for any n, as long as we continue to have $e_t^2(n) \ge e_t^1(n) = 0$ for all $t \ge T$, we will continue to have $e_t^{*2}(n) >> e_t^{*1}(n) = 0$ for all $T \le t < T' - 1$. Since the belief-revision dynamics (2.18) 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 $0 \le \tilde{n} < n$, as long as we have had $e_t^{*2}(\tilde{n}) >> 0$ for all $0 \le \tilde{n} < n$, 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 $T \le t < T' - 1$.

The associated reflective equilibrium outcomes are given by (A.1) 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 $e_t^2(n) \ge e_t^1(n) = 0$ for all $T \le t \le T' - 1$ and $\bar{\imath}_{SR} < 0$ implies that $x_t^2 >> x_t^1$ for all $T \le t \le T' - 1$.

Finally, let us consider reflective equilibrium in periods $0 \le t < T$. 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, $e_t^2 \ge e_t^1$. Then

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

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

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

DTBC – 796
Zero Lower Bound Risk and Long-Term Inflation in a Time Varying Economy
Benjamín García

DTBC – 795
An Analysis of the Impact of External Financial Risks on the Sovereign Risk Premium of Latin American Economies
Rodrigo Alfaro, Carlos Medel y Carola Moreno

DTBC – 794
Welfare Costs of Inflation and Imperfect Competition in a Monetary Search Model
Benjamín García

DTBC – 793
Measuring the Covariance Risk Consumer Debt Portfolios
Carlos Madeira

DTBC – 792
Reemplazo en Huelga en Países Miembros de la OCDE: Una Revisión de la Legislación Vigente
Elías Albagli, Claudia de la Huerta y Matías Tapia

DTBC – 791
Forecasting Chilean Inflation with the Hybrid New Keynesian Phillips Curve: Globalisation, Combination, and Accuracy
Carlos Medel

BANCO CENTRAL
DE CHILE