Forward Guidance and the Role of Central Bank Credibility under Heterogeneous Beliefs

Gavin Goy[1], Cars Hommes[1,3], and Kostas Mavromatis[2]

[1] *CeNDEF, University of Amsterdam*
[2] *MInt, University of Amsterdam*
[3] *Tinbergen Institute, Amsterdam*

This Version: June 7, 2016

Abstract

This paper studies the macroeconomic effects of central bank forward guidance when agents are boundedly rational and heterogeneous. In particular, we take a cashless New Keynesian model subject to an occasionally binding zero lower bound constraint on nominal interest rates and replace the representative agent with a more realistic population of boundedly rational agents. Private households can only form expectations for a finite horizon and use simple forecasting heuristics to forecast key macroeconomic variables of interest. Further, these agents switch endogenously between these heuristics according to a performance measure. Contrary, the central bank is assumed to use a bivariate VAR to form expectations, yet is unaware of the time-variation in the distribution of aggregate expectations. Besides the nominal interest rate, the central bank has two additional policy tools at its disposal: it can publish its own forecasts (Delphic) and commit, though only imperfectly credible, to a future path of nominal interest rates (Odyssean). We find that the combination of real-time learning and model misspecification of the central bank gives rise to policy mistakes that can result in periods of high inflation. Moreover, both Delphic and Odyssean forward guidance are useful tools in stabilizing the economy, although, when private households are limitedly forward-looking, the benefits of the latter alone are marginal. Delphic forward guidance, on the other hand, significantly lowers the likelihood of deflationary spirals, once the economy is locked in a liquidity trap.

JEL-Classification: tba
Keywords: tba

1 Introduction

In the aftermath of the financial crisis 2007-9 central banks in the developed world successively cut the nominal interest rate until reaching the zero lower bound (ZLB). With the economy still in need of further monetary accommodation, several central banks decided to adopt new and unconventional monetary policy tools, including special lending facilities, large scale asset purchase programs (also known as *quantitative easing*) and forward guidance. The latter received brisk attention in the academic debate.[1] Following Campbell et al. (2012) forward guidance can be divided into two types: *Delphic* and *Odyssean* guidance. Under Delphic forward guidance we mean the central bank's publication of its own projections of key macroeconomic variables. Odyssean forward guidance, on the other hand, can be interpreted as a central bank's commitment to temporarily deviate from its policy rule and not respond to rising inflation and growth, but instead keep rates close to zero for a longer period. Therefore, this policy can also be seen as a *lower-for-long* policy (Haberis et al., 2014). If the central bank succeeds in convincing the market that it can indeed tie its hands, market's forecasts for nominal interest rate will be lower, which in turn reduces longer term real interest rates and thus stimulates spending and economic activity.

However, the effectiveness of central bank forward guidance crucially hinges on the central bank's credibility.[2] Furthermore, there is a broad consensus among both economists and central bankers that credibility needs to be build up over time (Blinder, 1999). This raises the question whether it is reasonable to take credibility as given−as usually done in the literature−or whether doing so leads to severely biased policy implications. In fact, already King (2009) "calls for the development of models of the zero bound with imperfect credibility" which can address the issues arising from optimal monetary policy under commitment in the ZLB region. In line with those thoughts, we argue that a more natural approach is to think of credibility as being dependent on how well the central bank performed in achieving its targets, like, for instance, price stability and/or high employment. Therefore, we propose a model of bounded rationality which endogenizes central bank credibility to analyze the effects of forward guidance.

In this paper we study the macroeconomic effects of central bank forward guidance when agents are boundedly rational and heterogeneous. In particular, we take a cashless New Keynesian model subject to an occasionally binding zero lower bound constraint on nominal interest rates and replace the representative household with a more realistic population of boundedly rational households. In fact, bounded rationality is reflected in multiple dimensions. Firstly, we assume that private households can only form expectations for a *finite* horizon. Specifically, we take the Euler equation as a behavioral primitive governing private households' intertemporal consumption decision and iterate it for a finite number of periods as done in Branch et al. (2012) and Ferrero and Secchi (2011). Secondly, we assume that private households use *simple* forecasting heuristics to forecast key macroeconomic variables of interest. The individual's choice of which forecasting heuristic to use is endogenous and depends on the heuristic's past performance. Concerning the heuristics, we allow agents to use either a purely backward-looking adaptive learning rule or to follow the central bank's announcements, i.e. announced targets or projections. We will argue that the latter can be

[1] A growing literature has analyzed the effects of forward guidance in a New Keynesian framework. Carlstrom et al. (2012) and others conclude the effect of forward guidance to be implausibly large in this class of models; an observation later called the *forward guidance puzzle* by Del Negro et al. (2012). While Cochrane (2015) argues that it is the failure of the New Keynesian theory that causes this result, García-Schmidt and Woodford (2015), however, attribute it to unrealistic assumption of perfect foresight or rational expectations and show that once this assumption is relaxed, the basic New Keynesian model generates empirically realistic result.

[2] See Blinder (1999) for the importance of credibility in monetary policy.

seen as an endogenous measure of central bank credibility. Thus, our approach offers a natural way to relax the *ad hoc* assumption of exogenous credibility. Ultimately, this leads to heterogeneity in private sector expectations with a time-varying distribution.[3] Next to the private households, our model features a central bank, which is assumed to use a bivariate vector autoregression (VAR) model to form expectations. In fact, central banks allocate a large share of resources into forecasting and thus we believe the central bank should also be the agent in the economy with the most sophisticated forecasting technology. However, the central bank has only limited information about the structure of the economy and therefore ignores the time-variation in the distribution of private sector expectations by assuming time-invariant coefficients. Due to this misspecification in the central bank's perceived law of motion of the economy, an equilibrium can be regarded as a *restricted perception equilibrium* (RPE).[4] Furthermore, our equilibrium definition is also related to the concept of a *stochastic consistent expectations equilibrium* studied by Hommes and Sorger (1998), which is a particular type of RPE where the underlying model is nonlinear, yet agents use a linear model to form forecasts. In our case, the specific nonlinearity is the time-variation in the distribution of aggregate expectations, that affects the coefficients of the underlying law of motion. Ultimately, the presence of a central bank results in a second dimension of heterogeneity in expectations, namely between private sector and central bank. Hence, the model exhibits *structural heterogeneity* (Evans and Honkapohja, 2001, p.42), since we have heterogeneity in the expectation formation even in equilibrium. Concerning the feedback of agents' expectations, it is important to see that while private sector expectations influence the economy directly through the IS curve and the New Keynesian Philips curve, the central bank's forecasts will only enter indirectly through the nominal interest rate rule. Moreover, we show that in normal times, that is when the ZLB constraint is not binding, only the central bank's inflation expectation matter, while once the ZLB is reached and the central bank conducts forward guidance also its output gap expectations are important. This model feature is comparable to the Fed's approach to its dual mandate, as in the recent recession the Fed focused more on boosting employment and economic activity rather than controlling for medium-run inflation.

In this bounded rationality framework we analyze both the coordinating effect of the central bank's publication of own inflation and output gap projections as well as the macroeconomic effects of a lower-for-long policy. Both policies are of a particular interest as the central bank's credibility evolves endogenously in our model. Firstly, we find that the stimulating effect of an announced lower-for-long policy is considerably dampened, if only a fraction of households expect short-term interest rate to remain close to zero for the next periods. Further, neither inflation nor output gap suddenly jumps at the moment of the announcement. Thus, we provide a potential solution to the *forward guidance puzzle*. Secondly, we find that the ability of the central bank to publish its own forecasts of inflation and output gap can help to coordinate individuals' expectations, yet the degree depends on the accuracy of its forecasts. This is in line with Ferrero and Secchi (2009), who study the case of the Reserve Bank of New Zealand (RBNZ) that publishes its own interest rate projections since 1999, and show that market expectations of the short term interest rate respond

[3] The empirical evidence of heterogeneity in expectations is numerous. To name a few, Carroll (2003), Mankiw et al. (2003), Pfajfar (2009) and Pfajfar and Santoro (2010) provide empirical support for heterogeneity in expectations using survey data on inflation expectations. Furthermore, Hommes et al. (2005), Adam (2007), Hommes (2011), Pfajfar and Zakelj (2011) and Assenza et al. (2013) find evidence for heterogeneity in learning-to-forecast laboratory experiments with human subjects. Importantly, also the distribution of heterogeneity evolves over time in response to economic volatility (see Mankiw et al., 2003); a feature well captured by our heuristic switching model.

[4] For a detailed discussion (see Evans and Honkapohja, 2001, Chapter 3.6 and 13)

in a significant and consistent way to the unexpected component of the published path, even though the adjustment is not complete. That is, the public moves its expectations only partially in the direction of the announcement. Concerning the effectiveness of forward guidance, we find that both Delphic and Odyssean forward guidance are useful tools in stabilizing the economy. However, when private households are limitedly forward-looking, the benefits of the latter alone are modest. Delphic forward guidance, on the other hand, significantly lowers the likelihood of deflationary spirals, once the economy is locked in a liquidity trap. The understanding of this result crucially hinges on the role of the *expected* real interest rates. Although the commitment to low nominal rates helps in reducing the expected real interest rates, also households inflation expectations are important. In fact, if those are deflationary enough households may believe to remain in the ZLB region. If the central bank in such a situation manages to influence also inflation expectations, such that private households also believe future inflation to pick up again, the expected real interest rates fall further, thereby stimulating economic activity today. Lastly, we observe that the combination of real-time learning, model misspecification and imperfect credibility of the central bank gives rise to policy mistakes that can result in periods of high inflation.

We are not the first to analyze forward guidance under bounded rationality. García-Schmidt and Woodford (2015) relax the assumption of perfect foresight and allow for different levels of reflection. On the other hand, Ferrero and Secchi (2011); Cole (2015); Honkapohja and Mitra (2015) study the effects of forward guidance under learning.

Moreover, our paper picks up the issue of real-time data availability in monetary policy discussed by Orphanides (2001) as well as central bank learning (Honkapohja and Mitra, 2005; Aoki and Nikolov, 2006). In particular, Honkapohja and Mitra (2005) stress that heterogeneity by central banks and private households pose a new stability concern that needs to be taken into account in policy design.

Most closely, however, our paper relates to Ferrero and Secchi (2011), Haberis et al. (2014) and Honkapohja and Mitra (2015). Although Haberis et al. (2014) and arguably Ferrero and Secchi (2011) consider the case with imperfect credibility, all three papers treat central bank credibility as exogenous. More specifically, Ferrero and Secchi (2011) study the effects central bank publications of macroeconomic projections have on the dynamic properties of an economy in which agents are learning. Similar to our approach, Ferrero and Secchi (2011) also iterate the New Keynesian IS and Philips curve, however, their approach differs in multiple dimensions. In particular, the authors assume households to use recursive least squares learning, but do not impose a similar informational constraint on the side of the central bank. Concerning the central bank's announcements, Ferrero and Secchi (2011) assume that a fixed fraction of households fully incorporate the policy announcements, while others not.[5] Contrary, our paper proposes a realistic mechanism to endogenize the fraction of agents who believe these policy announcements by implementing the heuristic switching model into the framework. Lastly, the authors ignore the ZLB constraint on nominal interest rates, which we believe to be an important nonlinearity in the context of forward guidance.

In a similar vein, Honkapohja and Mitra (2015) examine the global dynamics of New Keynesian model under learning, when the central bank targets either price level or nominal GDP. The fact that both policy regimes make monetary policy history-dependent allows the authors to implement forward guidance announcements into their model. Honkapohja and Mitra (2015) show that when

[5] More precisely, the authors assume a representative agent that uses a weighted average of own and central bank predictions with fixed weights, which can also be interpreted as different types of agents in the economy.

the policy announcements are incorporated into the private agents' learning, the basin of attraction of the targeted steady state under price level targeting increases substantially, as compared to inflation targeting. However, the extreme effects found by Honkapohja and Mitra (2015) seem to be (at least partly) driven by the *ad hoc* assumption that the central bank is perfectly credible in achieving either the nominal GDP of price level target.

A different approach is taken by Haberis et al. (2014), who show that the macroeconomic effects of a transient interest rate peg can be significantly dampened once the peg is believed to be only imperfectly credible. In fact, the authors assume that the central bank reneges on its announcement of keeping nominal interest rates at zero for a finite number of periods. This decision of when to renege on the promise or not is not modeled explicitly but stochastic with exogenous probability. Ultimately, the private sector's expectations of endogenous variables are a convex combination of the endogenous variables resulting from no peg and those which are consistent with the interest rate peg, both under rational expectation. Quite intuitively, as the credibility of the peg increases, the effects of forward guidance tends imitate those backward-explosive dynamics outlined by Carlstrom et al. (2012), while as credibility declines the effects of the peg tend to cease away. Although, our model also allows for imperfect credibility, we endogenize the households' perceived credibility of the central bank, while assuming that the central bank honors its promise as long as expected inflation does not exceed the target by too much. Moreover, we deviate from the rational expectations hypothesis and allow for heterogeneity in expectations.

Therefore, we see the main contribution of our paper in shedding more light on the effectiveness of forward guidance when the central bank's credibility itself is endogenous and agents are heterogeneous and boundedly rational. From a more methodological view point, we are−to the best of our knowledge−the first to combine finite Euler equation learning and the heuristic switching model.

The paper continues as follows. In Section 2 we will present our behavioral New Keynesian DSGE model as well as the results for E-stability. We move on by introducing the ZLB constraint in Section 3, where we first discuss the model without guidance in Section 3.1 and then allow central bank to make use of forward guidance announcements in Section 3.2. Our numerical results are illustrated in Section 4. In Section 5 we challenge some of our main assumptions to analyze the robustness of our results. Finally, Section 6 concludes.

2 The Model

2.1 New Keynesian framework

We start out with the log-linearized version of a standard cashless New Keynesian model (NKM) in line with Galí (2009). The model features monopolistic competition in the goods market as well as stickiness in price setting and is described by the two equations

$$x_t = E_t x_{t+1} - \frac{1}{\sigma}\left[i_t - E_t \pi_{t+1}\right] + e_t \tag{1}$$

$$\pi_t = \beta E_t \pi_{t+1} + \kappa x_t + u_t \tag{2}$$

which pin down output gap x_t and inflation π_t, given a nominal interest rate i_t. The variables e_t and u_t represent a demand (or real interest rate) shock and a cost-push shock, respectively.

The parameters β and σ are the households' discount factor and the inverse of the intertemporal elasticity of substitution, while κ captures the degree of price rigidity. That is, $\kappa = \infty$ implies that there are no price rigidities (the flexible price case) and the Philips curve (2) becomes a vertical line, whereas $\kappa = 0$ corresponds to complete price rigidity with a horizontal Philips curve.[6]

Our model follows Hommes and Lustenhouwer (2015) and−different to the representative agent model outlined in Galí (2009)−assumes a continuum of heterogeneous household-firms that differ in the way expectations are formed. However, we relax the assumption that agents only form one period ahead expectations. To be specific, we assume that private agents form expectations over a longer, but finite horizon as done in Branch et al. (2012). This is in sharp contrast to the rational expectations literature in which agents have sufficient sophistication to solve an infinite horizon dynamic programming problem. Instead, finite horizon learning generalizes the two existing benchmarks in the literature, namely Euler equation learning, which assumes that consumption decisions are made to satisfy the one-step-ahead perceived Euler equation (Evans and Honkapohja, 2006; Honkapohja et al., 2012), and infinite horizon learning, in which consumption today is determined optimally from an infinite-horizon optimization problem with given beliefs (Marcet and Sargent, 1989). Thus, under *N-step Euler equation learning* we iterate both the IS curve (1) as well as the Philips curve (2) N times to find

$$x_t = \tilde{E}_t x_{t+N} - \frac{1}{\sigma}\tilde{E}_t \sum_{j=0}^{N-1} \left(i_{t+j} - \pi_{t+j+1} - \sigma e_{t+j}\right) \tag{3}$$

$$\pi_t = \beta^N \tilde{E}_t \pi_{t+N} + \tilde{E}_t \sum_{j=0}^{N-1} \beta^j \left(\kappa x_{t+j} + u_{t+j}\right) \tag{4}$$

where $\tilde{E}_t$ denotes the heterogeneous expectations operator. Note further that both sequences of expected shocks in Equations (3) and (4), respectively, reduce to e_t and u_t, if we assume that either shocks are only transitory, or agents do neither know nor learn about the autoregressive nature of the shocks. For now, let us assume the latter, i.e. both shocks follow an AR(1) process, but agents are not aware of this fact. The processes are given by

$$e_t = \rho_x e_{t-1} + \varepsilon_x \tag{5}$$

$$u_t = \rho_\pi u_{t-1} + \varepsilon_\pi \tag{6}$$

with ε_x and ε_π being mutually orthogonal white noises with variances σ_x^2 and σ_π^2, respectively.

To close the model, an interest rate rule needs to be specified. Specifically, we assume the central bank to set the nominal interest rate according to a simple expected contemporaneous Taylor-type rule to adequately represent the informational problem faced by the central bank. [7] Under this assumption, the central bank adjusts the nominal interest rate once it expects current period's inflation to deviate from its target. Formally, the effective nominal interest rate i_t set by the central bank is a function−which is described below−of the policy rate i_t^{mp} that follows the reaction function

$$i_t^{mp} = \max\{0, \bar{\pi} + \phi(\pi_t^{e,cb} - \bar{\pi})\} \tag{7}$$

[6] Hommes and Lustenhouwer (2015, Appendix A) derive the same equations that arise under a representative household with rational expectations for heterogeneous agents under the assumption of Euler equation learning, using properties of the heuristic switching model.

[7] See Orphanides (2001) for a detailed discussion of informational problems in monetary policy.

where the max-term represents the zero lower bound (ZLB) constraint the central bank faces. Further, $\pi_t^{e,cb}$ denotes the central bank's own prediction for current period inflation, made at the beginning of period t, that is, before endogenous aggregate outcomes are realized.

2.2 Private sector expectations

At this point we have to specify how expectations are formed. As mentioned above, our assumption on the timing of expectation formation is that private agents as well as the central bank do not observe current aggregate outcomes (i.e. endogenous variables), but only lagged realizations and form their expectations at the beginning of each period. Also, current period shocks are not observed. Thus, private agents and the central bank share the same information set.

Private households are assumed to use simple forecasting heuristics to form their expectations about key macroeconomic variables. Further, we let the private agents choose these heuristics endogenously out of a set of forecasting heuristics according to their relative performance in recent past. This idea goes back to Herbert Simon (1984) who proposed to model human decision making as a rational choice between simple forecasting heuristics. Formally, these ideas are well-captured in the heuristic switching model (HSM) proposed by Brock and Hommes (1997).[8] Let us denote the fraction of agents using a specific forecasting heuristic h out of the set of forecasting heuristic H at time t by $n_{h,t}$, which follows a logistic distribution of the form

$$n_{h,t} = \frac{\exp(\frac{b}{2}U_{h,t-1})}{\sum_H \exp(\frac{b}{2}U_{h,t-1})} \quad \text{for } h \in H \tag{8}$$

$$U_{h,t-1} = \rho U_{h,t-2} - (1-\rho)\sum_{z\in\mathcal{Z}} \left(z_{t-1} - \tilde{E}_{h,t-2}z_{t-1}\right)^2 \quad \text{with } \mathcal{Z} = \{x, \pi, i\} \tag{9}$$

where $U_{h,t}$ is the performance measure here defined as the negative sum of squared forecast errors. The parameter $b \in [0, \infty)$ is called the *intensity of choice* and it governs the sensitivity towards forecast errors, i.e. how fast agents switch to the optimal forecasting heuristic. In the special case $b = 0$ agents never switch their strategy such that all fractions will be constant and equal to $\frac{1}{H}$. Contrary, in the other extreme case of $b = \infty$, all agents will use the same optimal strategy in each period. The latter case is sometimes referred to as *neoclassical limit*, because it represents the highest degree of rationality. Additionally to Brock and Hommes (1997) we included memory in the fitness measure ρ. Aggregate or market expectations are then given by the weighted average of individual expectations $\tilde{E}_t z_{t+j} = \sum_{h\in H} n_{h,t}\tilde{E}_{h,t}z_{t+j}$ for $z_{t+j} \in \{x_{t+j}, \pi_{t+j}, i_{t+j}\}$.

In the rest of the paper we will confine ourselves to the case with $H = 2$. Precisely, we assume two types of agents, namely *adaptive learners* and *credibility believers*. Adaptive learners form their expectations for all variables (output gap, inflation and the nominal interest rate) using an adaptive expectations rule. In fact, this adaptive expectations rule can be derived from agents using steady state learning with a constant gain parameter when the exogenous shocks are *iid*. In

[8] The empirical evidence in favor of the heuristics switching model is compelling. For instance, Branch (2004, 2007) finds that survey data on inflation expectations are consistent with a dynamic choice between statistical predictor functions. Further, Anufriev and Hommes (2012*a*,*b*) fitted the heuristics switching model to the data of asset pricing learning-to-forecast experiments (see Hommes, 2011, for a survey of laboratory experiments) and found that already four simple heuristics explain most of the observations. Also, Hommes et al. (2005) argue that laboratory experiments with human subjects are well suited to discipline the class of individual heuristics that boundedly rational subjects may use in their decision making process. Lastly, Assenza et al. (2013) use the same heuristics switching model as in Anufriev and Hommes (2012*a*,*b*) and find that the simple heterogeneous expectations switching model also fits individual learning and aggregate outcomes in the standard New Keynesian macroeconomic setting.

other words, adaptive learners treat inflation as an *iid* process with an unknown mean, which they try to estimate by least squares.[9] On the other hand, credibility believers fully believe in the central bank's ability to achieve its target of price stability, that is inflation to be at target $\bar{\pi}$ and output gap to be equal to its steady state level. These expectations coincide with the rational expectation values in the absence of any other agent and no auto-correlated shocks (for detailed calculations see Section A in the Appendix). However−and this is a key novelty of our paper−credibility believers also adjust their forecasts if the central bank uses forward guidance. Specifically, we assume that the central bank can publish its own j-periods ahead projections for inflation and output gap (denoted as $\pi^{e,cb}_{t+j}$ and $x^{e,cb}_{t+j}$, respectively), which we interpret as *Delphic* forward guidance, and further can announce a future path for the nominal interest rate, which we regard as *Odyssean* type of forward guidance. More precisely, the Odyssean guidance is a conditional promise of the central bank to keep nominal interest rates at zero for a prolonged period of time, as long as the central bank's own projections of next period's inflation do not exceed its target $\bar{\pi}$ too much.[10] Therefore, let $\tilde{\pi}$ be a threshold level of inflation with $\tilde{\pi} > \bar{\pi}$, for which it holds that if $\pi^{e,cb}_t > \tilde{\pi}$ the central bank will revert back to its usual reaction function given by Equation (7).[11] Formally, we can write the interest rate policy that results from the Odyssean guidance as

$$i_t = \begin{cases} i^{mp}_t & \text{if} \quad i^{mp}_{t-j} > 0, \quad \forall j = 1, \ldots, q \quad \text{or} \quad \pi^{e,cb}_t > \tilde{\pi} \\ 0 & \text{else} \end{cases} \tag{10}$$

with q denoting the forward guidance horizon.[12]

In the following, we describe private households' forecasting heuristics for the output gap, inflation and the nominal interest rate. Let the expectations of all households using forecasting heuristic h for variable z_{t+j} at time t be denoted as $\tilde{E}_{h,t} z_{t+j}$. We then have a household's output gap expectations given by

$$\tilde{E}_{1,t} x_{t+j} = \begin{cases} x^{e,cb}_{t+j}, & \forall j = 1, \ldots, q \\ \bar{x}, & \forall j = q+1, \ldots, N \end{cases} \tag{11}$$

and

$$\tilde{E}_{2,t} x_{t+j} = \tilde{E}_{2,t-1} x_t + \omega\big(x_{t-1} - \tilde{E}_{2,t-1} x_t\big), \quad \forall j = 1, \ldots, N \tag{12}$$

respectively, where the first row in Equation (11) corresponds to the central bank's announcement of its own output gap projections $x^{e,cb}_{t+j}$ for the next q periods. How the central bank obtains

[9] See Evans et al. (2008) for a more detailed discussion.

[10] This assumption is in line with the Fed's announcement from December 12, 2012. The relevant part reads: "*[T]he Committee decided to keep the target range for the federal funds rate at 0 to 1/4 percent and currently anticipates that this exceptionally low range for the federal funds rate will be appropriate at least as long as ... inflation between one and two years ahead is projected to be no more than a half percentage point above the Committee's 2 percent longer-run goal, and longer-term inflation expectations continue to be well anchored*". While short term inflation expectations of the credibility believers can be influenced through the Delphic guidance statements, their longer term inflation expectations are anchored at the target in our model as long as $N > q$.

[11] Generally, it is possible to micro-found this threshold inflation level $\tilde{\pi}$ by formulating the central bank's trade-off between the loss of credibility when continuing the lower-for-long policy due to higher inflation versus the loss of credibility from reneging from the previous promise. King (2009), for instance, shows that both inflation and output gap are optimally higher under commitment, when the economy exits the ZLB.

[12] Hanson and Stein (2015) and Swanson and Williams (2014) argue that the Fed's forward guidance strategy operates with a roughly two-year horizon. In our model, this corresponds to $q = 8$ since one time period is a quarter.

its projections will be specified below. Moreover, note that in the absence of forward guidance ($q = 0$) credibility believers expect output gap to be at the rational expectations equilibrium (with a representative agent) every period. The parameter ω in the adaptive learning rule (12) is the so-called *gain parameter*, which we assume to be constant. For $\omega = 1$ these expectations reduce to *naive* expectations. Similar to the output gap expectations, we can can summarize the inflation expectations as

$$\tilde{E}_{1,t}\pi_{t+j} = \begin{cases} \pi^{e,cb}_{t+j}, & \forall j = 1, \dots, q \\ \bar{\pi}, & \forall j = q+1, \dots, N \end{cases} \tag{13}$$

and

$$\tilde{E}_{2,t}\pi_{t+j} = \tilde{E}_{2,t-1}\pi_t + \omega\big(\pi_{t-1} - \tilde{E}_{2,t-1}\pi_t\big), \quad \forall j = 1, \dots, N \tag{14}$$

respectively, where again the first row in Equation (13) corresponds to the central bank's announcement of its own inflation projections $\pi^{e,cb}_{t+j}$.

Next, recognize that, due to the forward-iteration of the IS curve (3), agents also have to form expectations about the nominal interest rate. For this reason we assume that agents are aware of the ZLB constraint and, moreover, know the functional form of central bank's reaction function in normal times. However, private agents use their own inflation expectations to determine the nominal interest rate that is consistent with their beliefs. Under these assumptions, the nominal interest rate expectations are given by

$$\tilde{E}_{1,t}i_{t+j} = \begin{cases} 0, & \forall j = 1, \dots, q \\ \bar{\pi}, & \forall j = q+1, \dots, N \end{cases} \tag{15}$$

and

$$\tilde{E}_{2,t}i_{t+j} = \max\{0, \bar{\pi} + \phi(\tilde{E}_{2,t}\pi_{t+j} - \bar{\pi})\}, \quad \forall j = 1, \dots, N \tag{16}$$

Similar as above, the first row of Equation (15) captures the forward guidance policy, more precisely, the lower-for-long policy (or Odyssean guidance). However, in the absence of such policies credibility believers expect the nominal interest rate to be at its target, which also equals $\bar{\pi}$ because we normalized the equilibrium real interest rate to zero. Equation (14) corresponds to the interest rate expectations of adaptive learners.

2.3 Central bank learning

Lastly, we specify how the central bank forms its forecasts. It seems to be a reasonable assumption that the central bank has the most sophisticated forecasting model, as central banks generally devote large amounts of resources on forecasting. For this reason, we assume that the central bank uses a first-order bivariate VAR model to estimate future inflation and output gap.[13] Formally,

[13] Since the central bank uses the VAR model primarily for forecasting, a parsimonious model−like a VAR(1)−seems appropriate.

the central bank estimates the following VAR model

$$
\begin{aligned}
x_t^{e,cb} &= a_{10} + a_{11}x_{t-1} + a_{12}\pi_{t-1} \\
\pi_t^{e,cb} &= a_{20} + a_{21}x_{t-1} + a_{22}\pi_{t-1}
\end{aligned}
$$

which can be simplified to

$$y_t^{e,cb} = A' w_{t-1} \tag{17}$$

where $y_t^{e,cb} = [x_t^{e,cb}, \pi_t^{e,cb}]'$ and $w_{t-1} = [1, x_{t-1}, \pi_{t-1}]'$. The matrix A is the corresponding 3×2 coefficient matrix, whose elements are updated each period using the following recursive least squares (RLS) algorithm:

$$A_t = A_{t-1} + \gamma_t R_t^{-1} w_{t-1} \big(y_t - A'_{t-1} w_{t-1}\big)' \tag{18}$$

$$R_t = R_{t-1} + \gamma_t \big(w_{t-1} w'_{t-1} - R_{t-1}\big) \tag{19}$$

where R_t is a moment matrix and γ_t is the so called *gain parameter*, which we assume to be decreasing and equal to $\gamma_t = \frac{1}{t}$. In the learning literature Equation (17) is also called the *Perceived Law of Motion* (PLM).[14] We assume *anticipated utility* behavior of the central bank. That is, the central bank believes that the parameter estimates will remain unchanged in the future and does not take into account the fact that it is likely to revise them subsequently. Hence, the central bank's behavior deviates from full rationality as the bank does not take account of the effects its decisions have on future learning and it ignores the period-by-period model misspecification (Sargent, 1999). Moreover, the heterogeneity in private sector's expectation creates time-variation in the coefficients of the economic model which the central bank ignores by using a time-invariant PLM. In doing so, the central bank effectively assumes homogeneity in agent expectations. Thus, the misspecified PLM cannot possibly converge to its rational expectations equilibrium (i.e. when assuming the central bank has to estimate the minimum state variable solution which includes the difference in fractions m_t), but will converge to an equilibrium in which the central bank's beliefs are confirmed.

We split the coefficient matrix A into a vector of constants A_0 and a 2×2 matrix A_1. Then, under the assumptions made above, the central bank's j-periods ahead forecasts $x_{t+j}^{e,cb}$ and $\pi_{t+j}^{e,cb}$ made in the beginning of period t evolve according to

$$y_{t+j}^{e,cb} = \big(I + A_{1,t-1} + A_{1,t-1}^2 + \ldots + A_{1,t-1}^{j-1}\big) A_{0,t-1} + A_{1,t-1}^j y_{t-1} \tag{20}$$

where I denotes a 2×2 identity matrix. These forecasts can be published by the central bank to influence agents' expectations. However, the fact that the central bank has imperfect knowledge about structure of the economy−i.e. uses a misspecified VAR model to form expectations (as it neglects the fact that the fractions of agents switch over time) as well as the imprecision in its parameter estimates−may lead to policy mistakes that affect the performance of the central bank as well as its credibility (Aoki and Nikolov, 2006). In the following section we discuss the expectational stability properties of the central bank's learning. We will see that, if the fraction of private households in the economy using a specific heuristic is fixed, the system under learning is

[14] We restrict the PLM to be time-invariant, because the central bank ignores heterogeneity in private sector expectations. If one relaxes this assumption, however, the matrix A_t in Equation (17) would then be allowed to vary over time. This is, for instance, captured by *constant gain learning* where agents assign a higher weight to the most recent information by discounting past data, or by *Kalman filtering*.

stable. However, allowing the fractions to be time-varying may prevent the coefficients to converge.

2.4 Expectational stability

To simplify the analytical analysis of the expectational stability of the central bank's VAR model we set the gain parameter ω in the adaptive learner's forecasting heuristic to $\omega = 1$. Hence, we consider the case with *naive* households. Further, before deriving the dynamic system, it is helpful to define the difference in fractions as:

$$m_{t+1} \equiv n_{1,t+1} - n_{2,t+1} = \frac{\exp\left(\frac{b}{2}U_{1,t}\right) - \exp\left(\frac{b}{2}U_{2,t}\right)}{\exp\left(\frac{b}{2}U_{1,t}\right) + \exp\left(\frac{b}{2}U_{2,t}\right)} = \tanh\left(\frac{b}{2}\Delta U_t\right) \tag{21}$$

with $\Delta U_t \equiv U_{1,t} - U_{2,t}$. So, the term m_{t+1} equals 1 when all households are credibility believers, and -1 when all households are naive. Note that we can regard this expression as an endogenous measure of the *central bank's credibility*, because a higher m_{t+1} indicates higher confidence of the population in the central bank achieving its targets. ΔU_t describes the difference in performance measures−which we defined as the sum of negative squared forecast errors−for both heuristics. Henceforth, we will use the term fractions and difference in fractions interchangeably for m_{t+1}. The stochastic dynamic system described by Equations (1) - (17) and (21) in normal times (i.e. with a slack ZLB constraint and without forward guidance announcements) is then given by

$$\begin{aligned} x_t =& \left[\frac{1-\beta}{\kappa}\left(\frac{1+m_t}{2}\right) + \frac{(N-1)\phi - N}{\sigma}\left(\frac{1-m_t}{2}\right) + \frac{\phi}{\sigma}\right]\bar{\pi} \\ &+ \left(\frac{1-m_t}{2}\right)x_{t-1} - \left[\frac{(N-1)\phi - N}{\sigma}\left(\frac{1-m_t}{2}\right)\right]\pi_{t-1} - \frac{\phi}{\sigma}\pi_t^{e,cb} \end{aligned} \tag{22}$$

$$\begin{aligned} \pi_t =& \left[\frac{1+m_t}{2} + \frac{\kappa\left[(N-1)\phi - N\right]}{\sigma}\left(\frac{1-m_t}{2}\right) + \frac{\kappa\phi}{\sigma}\right]\bar{\pi} \\ &+ \kappa\left(\frac{1-\beta^N}{1-\beta}\right)\frac{1-m_t}{2}x_{t-1} + \left[\left(\beta^N - \frac{\kappa\left[(N-1)\phi - N\right]}{\sigma}\right)\frac{1-m_t}{2}\right]\pi_{t-1} - \frac{\kappa\phi}{\sigma}\pi_t^{e,cb} \end{aligned} \tag{23}$$

or in matrix notation

$$y_t \equiv \begin{pmatrix} x_t \\ \pi_t \end{pmatrix} = \Lambda_0(m_t) + \Lambda_1(m_t)y_{t-1} + \Lambda_2 y_t^{e,cb} \tag{24}$$

where we omitted the random disturbance terms for simplicity.[15] This can be done without loss of generality, because shocks are not observed by agents and thus do not affect the mapping. The matrices $\Lambda_0(\cdot), \Lambda_1(\cdot)$ and Λ_2 are given by

$$\Lambda_0(m_t) = \begin{pmatrix} \left[\frac{1-\beta}{\kappa}\left(\frac{1+m_t}{2}\right) + \frac{(N-1)\phi-N}{\sigma}\left(\frac{1-m_t}{2}\right) + \frac{\phi}{\sigma}\right]\bar{\pi} \\ \left[\frac{1+m_t}{2} + \frac{\kappa\left[(N-1)\phi-N\right]}{\sigma}\left(\frac{1-m_t}{2}\right) + \frac{\kappa\phi}{\sigma}\right]\bar{\pi} \end{pmatrix}$$

$$\Lambda_1(m_t) = \begin{pmatrix} \frac{1-m_t}{2} & -\frac{(N-1)\phi-N}{\sigma}\left(\frac{1-m_t}{2}\right) \\ \kappa\left(\frac{1-\beta^N}{1-\beta}\right)\frac{1-m_t}{2} & \left(\beta^N - \frac{\kappa\left[(N-1)\phi-N\right]}{\sigma}\right)\frac{1-m_t}{2} \end{pmatrix} \quad \text{and} \quad \Lambda_2 = \begin{pmatrix} 0 & -\frac{\phi}{\sigma} \\ 0 & -\frac{\kappa\phi}{\sigma} \end{pmatrix}$$

[15] Explanatory footnote ... or add them

respectively. Substituting the central bank's expectations gives rise to the *Actual Law of Motion* (ALM):

$$\begin{aligned} y_t =& \Lambda_0(m_t) + \Lambda_1(m_t)y_{t-1} + \Lambda_2\big[A_0 + A_1 y_{t-1}\big] \\ =& \Lambda_0(m_t) + \Lambda_2 A_0 + \big[\Lambda_1(m_t) + \Lambda_2 A_1\big]y_{t-1} \end{aligned} \tag{25}$$

where I denotes the identity matrix of according size. The ALM can be seen as describing the stochastic process followed by the economy if the central bank's forecasts are made under the fixed rule given by the PLM. In other words, given both the central bank's forecasts and those of the private sector the economy attains a *temporary equilibrium*.

Now, let us stag both columns of the coefficient matrix A on top of each other to form a 6×1 vector denoted by $\mathbf{a}$ (formally $\text{vec}A = \mathbf{a}$) and define the T-map as the mapping from the PLM to the ALM, which describes the evolution of the RLS estimator $\mathbf{a}$.[16] This mapping can be written as

$$T(\mathbf{a}) = \quad T\begin{pmatrix} a_{10} \\ a_{11} \\ a_{12} \\ a_{20} \\ a_{21} \\ a_{22} \end{pmatrix} = \begin{pmatrix} \left(\frac{1-\beta}{\kappa}\left(\frac{1+m_t}{2}\right) + \frac{(N-1)\phi-N}{\sigma}\left(\frac{1-m_t}{2}\right) + \frac{\phi}{\sigma}\right)\bar{\pi} - \frac{\phi a_{20}}{\sigma} \\ \frac{1-m_t}{2} - \frac{\phi a_{21}}{\sigma} \\ -\frac{\left[(N-1)\phi-N\right]}{\sigma}\frac{1-m_t}{2} - \frac{\phi a_{22}}{\sigma} \\ \left(\frac{1+m_t}{2} + \frac{\kappa\left[(N-1)\phi-N\right]}{\sigma}\left(\frac{1-m_t}{2}\right) + \frac{\kappa\phi}{\sigma}\right)\bar{\pi} - \frac{\kappa\phi a_{20}}{\sigma} \\ \kappa\left(\frac{1-\beta^N}{1-\beta}\right)\frac{1-m_t}{2} - \frac{\kappa\phi a_{21}}{\sigma} \\ \left(\beta^N - \frac{\kappa\left[(N-1)\phi-N\right]}{\sigma}\right)\frac{1-m_t}{2} - \frac{\kappa\phi a_{22}}{\sigma} \end{pmatrix} \tag{26}$$

After straight forward algebra we find the fixed point to this mapping that satisfies $T(\mathbf{a}^*) = \mathbf{a}^*$ to

[16] When continuing in matrix notation introduced above, the T-map can alternatively be written as

$$T\begin{pmatrix} A_0 \\ A_1 \end{pmatrix} = \begin{pmatrix} \Lambda_0 + \Lambda_2 A_0 \\ \Lambda_1 + \Lambda_2 A_1 \end{pmatrix}$$

with unique fixed point (for any given m)

$$\begin{aligned} A_0^* &= (I - \Lambda_2)^{-1}\Lambda_0 \\ A_1^* &= (I - \Lambda_2)^{-1}\Lambda_1 \end{aligned}$$

The modified E-stability condition is then determined by the corresponding ordinary differential equations

$$\frac{dA_0}{d\tau} \equiv T_0(A_0) - A_0 = \Lambda_0 + \Lambda_2 A_0 - A_0$$

and

$$\frac{dA_1}{d\tau} \equiv T_1(A_1) - A_1 = \Lambda_1 + \Lambda_2 A_1 - A_1$$

and it follows that convergence under under least squares learning occurs if the eigenvalues of $\Lambda_2 - I$ have real parts less than one. Since the eigenvalues are -1 and $-(1 + \kappa\phi/\sigma)$ this is the case.

be dependent on the difference in fraction m and given by

$$a^*_{10}(m) =\left(\frac{1-\beta}{\kappa}(\frac{1+m}{2}) + \frac{(N-1)\phi - N}{\sigma}(\frac{1-m}{2}) + \frac{\phi}{\sigma}\right)\bar{\pi} - \frac{\phi a^*_{20}}{\sigma} \tag{27}$$

$$a^*_{11}(m) =\frac{1-m}{2} - \frac{\phi a^*_{21}}{\sigma} \tag{28}$$

$$a^*_{12}(m) = -\frac{\big[(N-1)\phi - N\big]}{\sigma}\frac{1-m}{2} - \frac{\phi a^*_{22}}{\sigma} \tag{29}$$

$$a^*_{20}(m) =\frac{\sigma}{\sigma + \kappa\phi}\left(\frac{1+m}{2} + \frac{\kappa\big[(N-1)\phi - N\big]}{\sigma}(\frac{1-m}{2}) + \frac{\kappa\phi}{\sigma}\right)\bar{\pi} \tag{30}$$

$$a^*_{21}(m) =\frac{\kappa\sigma}{\sigma + \kappa\phi}\left(\frac{1-\beta^N}{1-\beta}\right)\frac{1-m}{2} \tag{31}$$

$$a^*_{22}(m) =\frac{\sigma}{\sigma + \kappa\phi}\left(\beta^N - \frac{\kappa\big[(N-1)\phi - N\big]}{\sigma}\right)\frac{1-m}{2} \tag{32}$$

where the fixed points for a^*_{10}, a^*_{11} and a^*_{12} depend on those of a^*_{20}, a^*_{21} and a^*_{22}, respectively, which are defined by (30)-(32). Thus, as long as fractions are fixed we have a unique fixed point of the T-map for every $m \in [-1, 1]$. In Appendix B we derive conditions for the parameters such that the VAR model of the central bank is stationary for different fixed points (i.e. for different values of m). In general, a non-stationary VAR model produces forecasts with unbounded error variances over longer horizons. Also the VAR-based forecasts itself will increase without bounds. If, however, the VAR model is only employed to form short run forecasts, the consequences of using a non-stationary VAR model might be limited.

A more interesting property of this fixed point is that the central bank's VAR model is stationary if and only if the dynamic system described by equation (25) is stable, because the central bank's VAR model coincides with the VAR of the economy as long as fractions are fixed. This result is summarized in Proposition 1.

Proposition 1. *The central banks VAR model is stationary if and only if the dynamic system of the economy is (locally) stable.*

Proof. To see this simply substitute the fixed point of the T-map (30)-(32), which can be rewritten as $A^*_0 = (I - \Lambda_2)^{-1}\Lambda_0$ and $A^*_1 = (I - \Lambda_2)^{-1}\Lambda_1$, respectively, into (25) to find

$$y_t = \big[I + \Lambda_2(I - \Lambda_2)^{-1}\big]\Lambda_0 + \big[I + \Lambda_2(I - \Lambda_2)^{-1}\big]\Lambda_1 y_{t-1}$$

Now, it is easy to verify that this system is equivalent to

$$y_t = A^*_0 + A^*_1 y_{t-1}$$

Therefore, also the dynamic system of the economy can be written as a VAR model. Further, it follows that the VAR model is stationary if the roots of $(1 - a^*_{11}L)(1 - a^*_{22}L) - (a_{12}a_{21}L^2)$ lie outside the unit circle, which coincides with the stability requirement requirement of our economic model. □

To determine whether the estimated coefficients converge to their respective fixed point for any of the given steady states, consider the following differential equation

$$\frac{d}{d\tau}(\mathbf{a}) = T(\mathbf{a}) - \mathbf{a} \tag{33}$$

where τ describes notional or artificial time. Component-by-component we get

$$\frac{da_{10}}{d\tau} = \Big(\frac{1-\beta}{\kappa}(\frac{1+m}{2}) + \frac{(N-1)\phi - N}{\sigma}(\frac{1-m}{2}) + \frac{\phi}{\sigma}\Big)\bar{\pi} - \frac{\phi a_{20}}{\sigma} - a_{10} \tag{34}$$

$$\frac{da_{11}}{d\tau} = \frac{1-m}{2} - \frac{\phi a_{21}}{\sigma} - a_{11} \tag{35}$$

$$\frac{da_{12}}{d\tau} = -\frac{\big[(N-1)\phi - N\big]}{\sigma}\frac{1-m}{2} - \frac{\phi a_{22}}{\sigma} - a_{12} \tag{36}$$

$$\frac{da_{20}}{d\tau} = \Big(\frac{1+m}{2} + \frac{\kappa\big[(N-1)\phi - N\big]}{\sigma}(\frac{1-m}{2}) + \frac{\kappa\phi}{\sigma}\Big)\bar{\pi} - \Big(1 + \frac{\kappa\phi}{\sigma}\Big)a_{20} \tag{37}$$

$$\frac{da_{21}}{d\tau} = \Big(\frac{1-\beta^N}{1-\beta}\Big)\frac{1-m}{2} - \Big(1 + \frac{\kappa\phi}{\sigma}\Big)a_{21} \tag{38}$$

$$\frac{da_{22}}{d\tau} = \Big(\beta^N - \frac{\kappa\big[(N-1)\phi - N\big]}{\sigma}\Big)\frac{1-m}{2} - \Big(1 + \frac{\kappa\phi}{\sigma}\Big)a_{22} \tag{39}$$

Now, note that the system described by Equation (34)-(39) is linear and the equations for (a_{10}, a_{20}), (a_{11}, a_{21}) and (a_{12}, a_{22}) are independent. We rewrite Equation (33) to get $\frac{d}{d\tau}(a) = M_0 + M_1\mathbf{a}$ with M_1 being a corresponding 6×6 matrix given by[17]

$$M_1 = \begin{pmatrix} -1 & 0 & 0 & -\frac{\phi}{\sigma} & 0 & 0 \\ 0 & -1 & 0 & 0 & -\frac{\phi}{\sigma} & 0 \\ 0 & 0 & -1 & 0 & 0 & -\frac{\phi}{\sigma} \\ 0 & 0 & 0 & -(1+\frac{\kappa\phi}{\sigma}) & 0 & 0 \\ 0 & 0 & 0 & 0 & -(1+\frac{\kappa\phi}{\sigma}) & 0 \\ 0 & 0 & 0 & 0 & 0 & -(1+\frac{\kappa\phi}{\sigma}) \end{pmatrix}$$

The 6×1 vector M_0 depends in general on the variable m_t and is therefore time-variant. Yet, since m_t is bounded, so will be M_0. Further, the matrix M_1 is independent of m_t with three eigenvalues taking on the value -1 and the other three being equal to $-(1 + \frac{\kappa\phi}{\sigma})$. Since all parameters are positive, all eigenvalues are negative and thus the differential equation is locally stable. In other words, the coefficients of the central bank's VAR model will always converge, if m converges, because the fixed point itself depends on the difference in fractions m.[18] Contrary, if we assume fixed fractions, the model becomes linear (except for the nonlinearity arising from the ZLB constraint) and Equations (27)-(32) depict a unique and (asymptotically) stable fixed point of the T-map. This *E-stability* result is summarized in Proposition 2.

Proposition 2. *Assume the fractions of households are constant, i.e. $m_t = m$, then the central bank learns the true model given in Equation (25). Under time-varying fractions, however, the true model features time-varying coefficients and the central bank learns the correct model if and only if the fractions converge. Otherwise, a restricted perception equilibrium is attained.*

Assuming the difference in fractions of private households m_t is constant and equal to zero (i.e. equal fractions), Figure 1 illustrates the convergence of the VAR coefficients to their respective fixed points, where we use our benchmark calibration as outlined in Table 1.

Generally, the law of motion of the model economy without binding ZLB constraint is a convex combination of the constant predictors stemming from the credibility believers and a system of

[17] Alternatively, we can look at the stability of the three subsystems (a_{10}, a_{20}), (a_{11}, a_{21}) and (a_{12}, a_{22}) as in Honkapohja and Mitra (2005), using the results of Honkapohja and Mitra (2006).

[18] As mentioned earlier, this equilibrium can be interpreted as a *restricted perception equilibrium*.

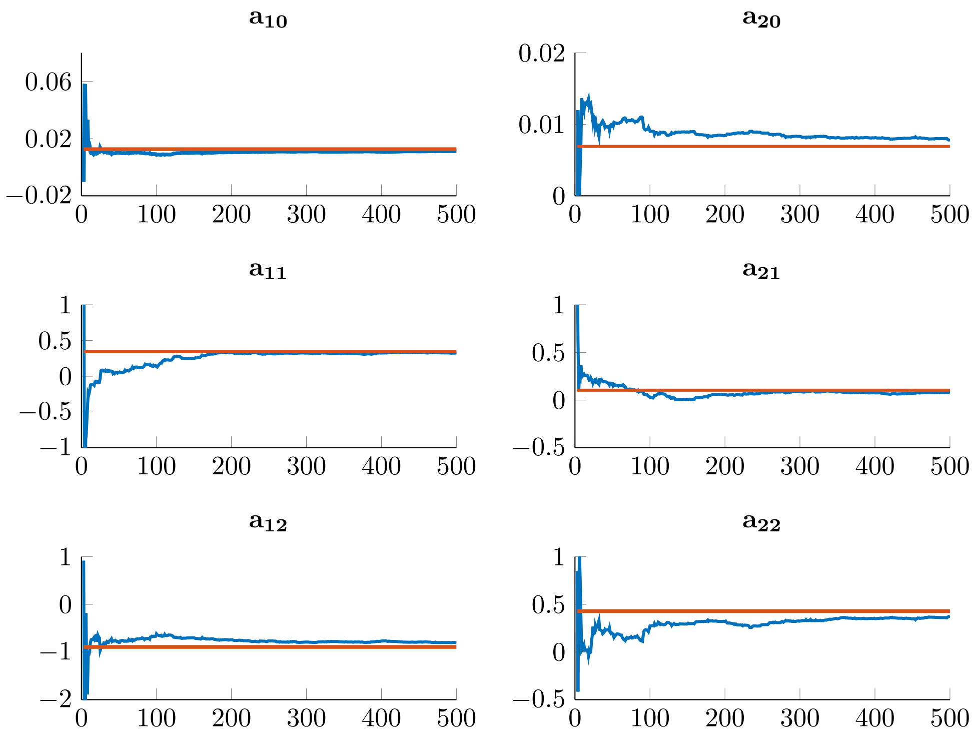

Figure 1: Convergence of VAR(1) coefficients as given in Equation (17) under equal and time-invariant fractions of both heuristics.

first order difference equations, stemming from both the central bank and the naive agents, with the specific weight attached to each of them being governed by the different fractions. Therefore, the lower m the larger will be the impact of the lagged variables and hence the larger will be the likelihood of the economic system to be unstable. This observation follows from the eigenvalues of the dynamic system. Both are zero if all private households are credibility believers and increase (in absolute value) in the fraction of naive households (see Appendix C for detailed calculations).

3 The model in the ZLB region

Given our specification of the interest rate rule in Equation (7), the central bank will lower the nominal interest rate once it expects current period's inflation $\pi_t^{e,cb}$ to be lower than the inflation target $\bar{\pi}$. Moreover, if the central bank expects inflation to be lower than $(1-\phi^{-1})\bar{\pi}$, the zero lower bound constraint will be binding and the central bank's conventional policy measures to stimulate the economy are exhausted. In this section we therefore analyze the effects of central bank communication, in particular the commitment to a *lower-for-long* policy and the publication of own inflation and output gap projections, to lift the economy out of the liquidity trap. Especially the first policy tool leads to mixed theoretical results. When agents have perfect foresight, the effects of forward guidance are highly overestimated (see Carlstrom et al. (2012) and García-Schmidt and Woodford (2015)) which can be partly attributed to the fact that the central bank is assumed to enjoy perfect credibility.[19] Contrary, when agents are learning (and thus are backward-looking) the

[19] Haberis et al. (2014) show that the effects of forward guidance can be significantly dampened, if agents place a non-zero probability on the event of the policymaker deviating from the lower-for-long commitment and reverting to his policy rule earlier than announced. However, credibility is treated exogenously and based on the time-inconsistency problem faced by the policymaker.

central bank has no credibility and announcement of such policies should have no immediate effects at the time of the announcement. Thus, both assumptions on how agents form their expectation can be interpreted as limiting cases. To fill this gap, this section proposes a model in which agents are forward-looking for a finite number of periods, and in which central bank credibility evolves endogenously depending on how well the central bank performed in achieving price stability.

3.1 The model without forward guidance

As mentioned above the ZLB constraint on nominal interest rates will be binding if the central bank's expectations for current inflation are pessimistic enough, i.e. satisfy

$$\pi_t^{e,cb} \leq (1 - \phi^{-1})\bar{\pi} \tag{40}$$

In the following we will refer to all (x_{t-1}, π_{t-1}) combinations that satisfy Equation (40) as the *zero lower bound region* or simply *ZLB region*. Consequently, the nominal interest rate i_t will be set to zero. To derive the economic model in the ZLB region without forward guidance, however, we further have to distinguish between two cases that depend on what naive agents expect in the near future to happen. To be precise, the naive agents' expectation rule in Equation (16) states that naive agents either expect the ZLB to be binding also in the next N periods, or they expect the economy leaving the liquidity trap already by the next period. Which of the two scenarios takes place crucially depends on period $t-1$ inflation. In fact, if π_{t-1} is sufficiently negative, i.e. satisfies

$$\pi_{t-1} \leq (1 - \phi^{-1})\bar{\pi} \tag{41}$$

then naive agents expect the ZLB to be binding for the next N periods. It is important to see that Equation (41)−other than Equation (40)−depends solely on period $t-1$ inflation. The economic model in the ZLB region without forward guidance is then given by

$$\begin{aligned} x_t =& \left[\left(\frac{1-\beta}{\kappa} + \frac{1}{\sigma}\right)\frac{1+m_t}{2} + \frac{(N-1)(\phi-1)\iota}{\sigma}\left(\frac{1-m_t}{2}\right)\right]\bar{\pi} + \left(\frac{1-m_t}{2}\right)x_{t-1} \\ &- \left[\frac{(N-1)\iota\phi - N}{\sigma}\left(\frac{1-m_t}{2}\right)\right]\pi_{t-1} \end{aligned} \tag{42}$$

$$\begin{aligned} \pi_t =& \left[\left(1 + \frac{\kappa}{\sigma}\right)\frac{1+m_t}{2} + \frac{\kappa(N-1)(\phi-1)\iota}{\sigma}\left(\frac{1-m_t}{2}\right)\right]\bar{\pi} + \kappa\left(\frac{1-\beta^N}{1-\beta}\right)\frac{1-m_t}{2}x_{t-1} \\ &+ \left[\left(\beta^N - \frac{\kappa\big[(N-1)\iota\phi - N\big]}{\sigma}\right)\frac{1-m_t}{2}\right]\pi_{t-1} \end{aligned} \tag{43}$$

with ι being an indicator function taking on value 0 if (41) holds and 1 otherwise.

Equations (42) and (43) illustrate that−conditional on Equation (40) to hold−the central bank's own forecasts do not enter the model equations, and hence there will also be no feedback from central bank's forecasts to the true realizations. It is therefore that we are not able to discuss E-stability properties in this particular case. In fact, as long as inflation satisfies Equation (40), Equations (42) and (43) represent an independent data generating process that is estimated by the central bank through RLS. However, given time-varying fractions, the central bank's model misspecification persists. As a result, it holds (again) that the central bank's estimated coefficients converge for a sufficiently large number of observations if and only if the difference in fractions converged.[20]

Next, we briefly consider the steady states of the dynamic system in the ZLB region. For this purpose we will distinguish between i) constant fractions, i.e. $m_t = m, \forall t$ and ii) time-varying fractions. The crucial difference is that, while under time-varying fractions the steady state difference in fractions m^* must not exceed zero, this is not a necessary condition under fixed fractions. To see this, recognize that adaptive learners (and thus naive households) always predict correctly in steady state and since, under time-varying fractions, households switch to the best performing heuristic(s), it must be that $m^* \leq 0$. Then, it becomes evident from Equation (3) and (4) that a unique ZLB steady state exists, featuring $x^* = \pi^* = i^* = 0$ and $m = -1$ if Equation (40) is satisfied.[21] The latter condition reduces to $1 - \phi^{-1} < 0$ given a positive inflation target $\bar{\pi} > 0$ and $Ez^* = (z^{e,cb})^* = z^* = 0$ with $z \in \{x, \pi, i\}$, which holds as long as the Taylor principle is adhered, i.e. the central bank should respond to a rise in inflation by a more than one-to-one increase in the nominal interest rate ($\phi > 1$). Contrary, if we fix the difference in fractions to any arbitrary $m^* \in (-1, 1)$, we can find a critical value for the difference in fraction $\hat{m}$ for which ZLB steady state only just exist. In fact, under constant fractions the model may feature only a single steady state with the share of credibility believers determining whether condition (40) is satisfied, or not. In other words, if the fraction of credibility believers is large enough there exists only one steady state outside the ZLB region, while for lower shares two steady states exist, one inside and one outside the ZLB region. These results are collected in the following Proposition. Its proof is left for the Appendix D.

Proposition 3. *Suppose $\phi > 1$ and $\bar{\pi} > 0$. Then the ZLB steady state exists*

i. for constant fractions if and only if $m \in [-1, \tilde{m}_1]$. Where $\tilde{m}_1$ is given by

$$\tilde{m}_1 = \frac{-\alpha_2 + \sqrt{\alpha_2^2 - 4\alpha_1\alpha_3}}{2\alpha_1}$$

ii. for time-varying fractions and features $\pi^* = x^* = i^* = 0$ with $m^* = -1$.

3.2 The model with forward guidance

In this section we allow the central bank to i) publish its own forecasts for inflation and the output gap and ii) to commit to a future interest rate path. Specifically, for the latter policy, we considers the commitment to keep nominal interest rates at zero for a prolonged period of time. The horizon for both policies be q periods. However, both these communicational policy measures are only picked up by the credibility believers.[22] Now, to illustrate the role of central bank publications in our behavioral model, consider the following scenario presented in Figure 2. In period 93 inflation was significantly below the central bank's target such that the central bank's inflation projections for the following period demand low nominal interest rates. Since, however, the central bank is constraint by the ZLB, it cannot lower rates sufficiently to provide the needed stimulus. In such a scenario, backward-looking households, that is those with adaptive or

[20] Note that in general $Corr(x_t, \pi_t) \neq 0$, because of the simultaneity laid out in Equations (3) and (4) (i.e. demand shocks will also affect inflation through the output gap). Nonetheless, OLS estimates remain unbiased.

[21] Actually, for intensities of choice less than infinity cycles of order-k, with $k \in \mathbb{N}$, might (theoretically) be possible, however, were not encountered in our simulation exercise in Section 4.2.

[22] We abstract from any noise in the communication channel of the central bank and leave this discussion for future research. There is no doubt that including both sender as well as receiver noise into the framework could generate interesting insights.

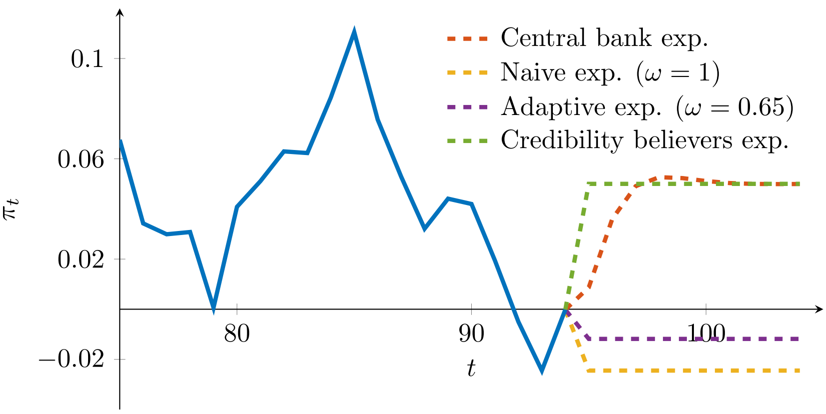

Figure 2: Different paths of (conditional) inflation expectations after hitting the ZLB in period 93. Benchmark calibration.

naive expectations, are pessimistic about the future development of inflation. On the other hand, credibility believers expect appropriate monetary policy to quickly lift inflation back to the central bank's target. While the latter arguably appears overly optimistic in such a scenario, the purely backward-looking expectations might be too pessimistic. However, if the fractions of adaptive learners is large enough, their deflationary expectations can potentially be self-fulfilling, such that adaptive learners could gain even further momentum under the dynamics of the heuristic switching and eventually inducing a deflationary spiral. To prevent this possibility, the central bank can try to influence the public's expectations by publishing it's own forecasts. In particular, the mechanism goes via the beliefs of the credibility believers, who adopt the central bank's forecasts and thus expect an arguably more reasonable path for inflation and output gap. Moreover, given the higher sophistication of the central bank's and thus leads to an increase in central bank credibility.

For simplicity assume that the forward guidance horizon equals the forward-looking horizon of the agents, i.e. $q = N$.[23]

3.2.1 Model with general N

Formally, the economic system in the ZLB region and with forward guidance is given by

$$\begin{aligned} x_t =& \frac{1-m_t}{2} x_{t-1} - \frac{(N-1)\iota\phi - N}{\sigma}\Big(\frac{1-m_t}{2}\Big)\pi_{t-1} + \frac{(N-1)(\phi-1)\iota}{\sigma}\Big(\frac{1-m_t}{2}\Big)\bar{\pi} \\ &+ \frac{1+m_t}{2} x^{e,cb}_{t+N} + \frac{1}{\sigma}\Big(\frac{1+m_t}{2}\Big)\sum_{j=1}^{N} \pi^{e,cb}_{t+j} \end{aligned} \tag{44}$$

$$\begin{aligned} \pi_t =& \frac{\kappa(2-\beta-\beta^N)}{1-\beta}\Big(\frac{1-m_t}{2}\Big)x_{t-1} + \frac{\kappa\iota(N-1)(\phi-1)}{\sigma}\Big(\frac{1-m_t}{2}\Big)\bar{\pi} \\ &+ \bigg(\beta^N - \frac{\kappa[(N-1)\iota\phi - N]}{\sigma}\bigg)\Big(\frac{1-m_t}{2}\Big)\pi_{t-1} + \kappa\Big(\frac{1+m_t}{2}\Big)\bigg(\sum_{j=1}^{N-1}\beta^j x^{e,cb}_{t+j} + x^{e,cb}_{t+N}\bigg) \\ &+ \frac{1+m_t}{2}\bigg[\frac{\kappa}{\sigma}\sum_{j=1}^{N-1}\pi^{e,cb}_{t+j} + (\beta^N + \frac{\kappa}{\sigma})\pi^{e,cb}_{t+N}\bigg] \end{aligned} \tag{45}$$

[23] We relax this assumption in Section 5.2. Contrary to the case with $q = N$, it turns out that for $q < N$, the ZLB steady state features an equilibrium difference in fractions of $m^* = -1$, instead of $m^* = 0$.

which is again conditional on Equation (40), i.e. central bank's expectations of period-t inflation must be deteriorated enough, as well as naive households' expectations, which depend on (41). Thus, ι is again the indicator function differentiating between the two cases of naive household's expectations. Rewriting Equations (44) and (45) in matrix notation, we get the following expectation-feedback system

$$y_t = \Lambda_0(m_t) + \Lambda_1(m_t)\begin{pmatrix} x_{t-1} \\ \pi_{t-1} \end{pmatrix} + \Lambda_2(m_t)\begin{pmatrix} x_{t+1}^{e,cb} \\ \pi_{t+1}^{e,cb} \end{pmatrix} \ldots + \Lambda_{N+1}(m_t)\begin{pmatrix} x_{t+N}^{e,cb} \\ \pi_{t+N}^{e,cb} \end{pmatrix} \tag{46}$$

where the coefficient matrices Λ_j are given below. Note, that these matrices generally depend on the difference in fractions m_t.

$$\Lambda_0(m_t) = \begin{pmatrix} \frac{(N-1)(\phi-1)\iota}{\sigma}\left(\frac{1-m_t}{2}\right)\bar{\pi} \\ \frac{\kappa\iota(N-1)(\phi-1)}{\sigma}\left(\frac{1-m_t}{2}\right)\bar{\pi} \end{pmatrix}, \qquad \Lambda_1(m_t) = \begin{pmatrix} \frac{1-m_t}{2} & -\frac{(N-1)\iota\phi-N}{\sigma}\left(\frac{1-m_t}{2}\right) \\ \kappa\left(\frac{2-\beta-\beta^N}{1-\beta}\right)\frac{1-m_t}{2} & \left[\beta^N - \frac{\kappa[(N-1)\iota\phi-N]}{\sigma}\right]\frac{1-m_t}{2} \end{pmatrix}$$

$$\Lambda_2(m_t) = \begin{pmatrix} 0 & \frac{1+m_t}{2\sigma} \\ \kappa\beta\left(\frac{1+m_t}{2}\right) & \frac{\kappa}{\sigma}\left(\frac{1+m_t}{2}\right) \end{pmatrix}, \qquad \Lambda_3(m_t) = \begin{pmatrix} 0 & \frac{1+m_t}{2\sigma} \\ \kappa\beta^2\left(\frac{1+m_t}{2}\right) & \frac{\kappa}{\sigma}\left(\frac{1+m_t}{2}\right) \end{pmatrix}$$

$$\ldots \qquad\qquad \ldots$$

$$\Lambda_N(m_t) = \begin{pmatrix} 0 & \frac{1+m_t}{2\sigma} \\ \kappa\beta^{N-1}\left(\frac{1+m_t}{2}\right) & \frac{\kappa}{\sigma}\left(\frac{1+m_t}{2}\right) \end{pmatrix}, \quad \Lambda_{N+1}(m_t) = \begin{pmatrix} \frac{1+m_t}{2} & \frac{1+m_t}{2\sigma} \\ \kappa\left(\frac{1+m_t}{2}\right) & \left(\beta^N + \frac{\kappa}{\sigma}\right)\frac{1+m_t}{2} \end{pmatrix}$$

Similar as in the exercise above we find that the matrices Λ_j are bounded given that m_t is bounded. In fact, all matrices Λ_j for $j > 1$ reduce to the null matrix if all private households are naive (i.e. $m_t = -1$) and we are back in the case of Section 3 without forward guidance. Intuitively, the central bank's *communication channel* ultimately depends on the fraction of credibility believers: if no household listens to the central bank's announcements, the effects of forward guidance are nil.

Next, let us iterate the central bank's PLM (17) forward and rewrite it as

$$y_{t+j}^{e,cb} = \bar{A}_0 + A_1^{j+1}(y_{t-1} - \bar{A}_0), \quad j = 0, \ldots q \quad \text{with} \quad \bar{A}_0 = (I - A_1)^{-1}A_0 \tag{47}$$

where $\bar{A}_0$ denotes a vector collecting the unconditional means. Substituting the central bank's forecasts into (46) we find the ALM in the ZLB region under forward guidance to be

$$\begin{aligned} y_t = \Lambda_0(m_t) + \Lambda_1(m_t)y_{t-1} + \Lambda_2(m_t)&\left[\bar{A}_0 + A_1^2(y_{t-1} - \bar{A}_0)\right] \\ + \Lambda_3(m_t)&\left[\bar{A}_0 + A_1^3(y_{t-1} - \bar{A}_0)\right] \\ &\vdots \\ + \Lambda_{N+1}(m_t)&\left[\bar{A}_0 + A_1^{N+1}(y_{t-1} - \bar{A}_0)\right] \end{aligned} \tag{48}$$

or simply

$$y_t = \left[\Lambda_0(m_t) + \sum_{j=1}^{N} \Lambda_{j+1}(m_t)\big(I - A_1^{j+1}\big)\bar{A}_0\right] + \left[\Lambda_1(m_t) + \sum_{j=1}^{N} \Lambda_{j+1}(m_t) A_1^{j+1}\right] y_{t-1} \tag{49}$$

Before turning to the E-stability conditions let us briefly analyze the steady states of the system (44) and (45). As in the Section 3 we again distinguish between constant and time-varying fractions. It turns out that under the forward guidance announcements considered the ZLB steady state exists for all fractions $m \in [-1, 1]$. Too see this, recognize that for $N \leq q$ the expectations of credibility believers coincide with those of the central bank. Therefore, the equilibrium with $\pi^* = x^* = i^* = 0$ can be supported by all values of m. For the same reason, i.e. the fact that both types of private households (naive and credibility believers) do not make any prediction errors in this steady state, the equilibrium difference in fractions m^* is necessarily equal to zero (equal fractions) once fractions are allowed to vary over time. This result is summarized in the following Proposition. Its proof is left for the Appendix E.

Proposition 4. *Suppose $\phi > 1$ and $\bar{\pi} > 0$. Then, given $N \leq q$, the ZLB steady state under forward guidance exists*

i. for all $m \in [-1, 1]$ if fractions are constant

ii. and for $m^* = 0$ if fractions are time-varying.

In both cases, steady state levels of inflation, output gap and the nominal interest rate are given by $\pi^* = x^* = i^* = 0$. If $N > q$, the results of Proposition 3 apply.

Moreover, the stability of the ZLB steady state under forward guidance with time-varying fractions ($m^* = 0$) is governed by the eigenvalues of the matrix $\Lambda_1(m^*) + \sum_{j=1}^{N} \Lambda_{j+1}(m^*)(A_1^*)^{j+1}$. In particular, if both eigenvalues lie inside the unit circle, the steady state is locally stable. However, the important question is whether this steady state is also learnable, because A_1^* shows up in the matrix equation. In what follows we will determine conditions for the learnability of the ZLB steady state under forward guidance. For notational convenience we suppress the dependence of $\Lambda_j(m_t)$ on the difference in fractions $\forall j = 0, \cdots, N+1$. We start with the mapping, T, from the PLM (17) to (49), which is given by

$$T\begin{pmatrix} A_0 \\ A_1 \end{pmatrix} = \begin{pmatrix} \Lambda_0 + \sum_{j=1}^{N} \Lambda_{j+1}\big(I - A_1^{j+1}\big)\bar{A}_0 \\ \Lambda_1 + \sum_{j=1}^{N} \Lambda_{j+1} A_1^{j+1} \end{pmatrix} \tag{50}$$

so that any fixed point (A_0^*, A_1^*) must necessarily satisfy

$$A_0^* = \left[I - \sum_{j=1}^{N} \Lambda_{j+1}\big(I - (A_1^*)^{j+1}\big)\big(I - A_1^*\big)^{-1}\right]^{-1} \Lambda_0 \tag{51}$$

$$\Lambda_1 + \sum_{j=1}^{N} \Lambda_{j+1}(A_1^*)^{j+1} - A_1^* = 0 \tag{52}$$

Then, for the ZLB steady state to be E-stable the differential equation $\frac{d}{d\tau}(A_0, A_1) = T(A_0, A_1) - (A_0, A_1)$ must have local asymptotic stability at the equilibrium (see Chapter 10 in Evans and Honkapohja, 2001). In other words, two conditions must be satisfied. First, the eigenvalues of

$DT_1(A_1^*)$ must have real parts less than unity, and second, provided that A_1 converges to A_1^*, also $DT_0(A_0^*, A_1^*)$ must have eigenvalues with real parts less than one. This result is summarized in Proposition 5.

Proposition 5. *Assume the T-map has a fixed point (A_0^*, A_1^*), i.e. in particular Equation (52) has a solution. Then the ZLB steady state is E-stable if*

i. all eigenvalues of $DT_1(A_0^*, A_1^*)$ have real parts less than one, and

ii. provided that A_1 converges to A_1^*, all eigenvalues of $DT_0(A_0^*, A_1^*)$ have real parts less than one

Since (52) is not solvable analytically we have to relied on numerical methods, however, no fixed point could be found.[24]

Next, we present the most simple case with $N = 2$ in the following section for illustrative purpose and to gain more insights in the E-stability condition.

3.2.2 Simplified model with $N = 2$

To analytically study the learnability properties and to better characterize the effects of the forward guidance policy of the central bank we abstract from some features in this section. In particular, we set the forward-looking horizon to the smallest possible, which is $N = 2$. The ALM then reduces to

$$y_t = \Lambda_0 + \big[\Lambda_2(I - A_1^2) + \Lambda_3(I - A_1^3)\big]\bar{A}_0 + \big[\Lambda_1 + \Lambda_2 A_1^2 + \Lambda_3 A_1^3\big]y_{t-1} \tag{53}$$

and, consequently, the T-map is given by

$$T\begin{pmatrix} A_0 \\ A_1 \end{pmatrix} = \begin{pmatrix} \Lambda_0 + \big[\Lambda_2\big(I - A_1^2\big) + \Lambda_3(I - A_1^3)\big](I - A_1)^{-1}A_0 \\ \Lambda_1 + \Lambda_2 A_1^2 + \Lambda_3 A_1^3 \end{pmatrix} \tag{54}$$

where we substituted $(I - A_1)^{-1}A_0$ back for $\bar{A}_0$. It immediately follows that a fixed point (A_0^*, A_1^*) to this mapping satisfies

$$A_0^* = \Big[I - \big[\Lambda_2\big(I - (A_1^*)^2\big) + \Lambda_3\big(I - (A_1^*)^3\big)\big]\big(I - A_1\big)^{-1}\Big]^{-1}\Lambda_0 \tag{55}$$

$$\Lambda_1 + \Lambda_2(A_1^*)^2 + \Lambda_3(A_1^*)^3 + A_1^* = 0 \tag{56}$$

Importantly, recognize that in the ZLB steady state we have that $\Lambda_0 = [0, 0]'$, since $\iota = 0$, and therefore A_0^* reduces to a zero vector. For Equation (56), however, no closed form solution exists. Relying on numerical methods we find that no solution exists to (56) under our baseline calibration.[25] Nonetheless we derive E-stability conditions below for completeness. These are, again, determined by the matrix differential equation

$$\frac{d}{d\tau}\big(A_0, A_1\big) = T\begin{pmatrix} A_0 \\ A_1 \end{pmatrix} - \begin{pmatrix} A_0 \\ A_1 \end{pmatrix}$$

[24] Using matlab's fsolve function, we search for solutions to this equation on a grid for N and m. We restricted our attention to $N \in \{2, \dots, 12\}$ and $m_t \in (-1, 1]$. Neither for the baseline calibration with $N = 4$, nor for other values of $N \in 1, \cdots, 12$ a fixed point exists.

[25] See Footnote 24.

First, let us focus on the mapping of A_1, because it is independent of A_0. We then have

$$\frac{d}{d\tau}(A_1) = \Lambda_1 + \Lambda_2 A_1^2 + \Lambda_3 A_1^3 - A_1 \equiv T_1(A_1) - A_1 \tag{57}$$

We say a particular solution A_1^* is E-stable if the fixed point of the differential Equation (57) is locally asymptotically stable at this point. To compute stability conditions, this equation has to be vectorized as done in Section 2.4 using the vec operator. Applying the rule for matrix differentials to Equation (57) we get

$$dT_1(A_1) = \Lambda_2(dA_1)A_1 + \Lambda_2 A_1(dA_1)I + \Lambda_3(dA_1)A_1^2 + \Lambda_3 A_1^2(dA_1)I \tag{58}$$

Then, the Jacobian of $\text{vec}T_1(A_1)$ is $DT_1(A_1) = \partial\text{vec}T_1(A_1)/\partial(\text{vec}A_1)'$. Using the rules $d\text{vec}A_1 = \text{vec}dA_1$ and $\text{vec}ABC = (C' \otimes A)\text{vec}B$ we find

$$DT_1(A_1) = A_1' \otimes \Lambda_2 + \left(A_1^2\right)' \otimes \Lambda_3 + I \otimes \left(\Lambda_2 A_1 + \Lambda_3 A_1^2\right) \tag{59}$$

and thus, the differential equation for A_1 is locally stable at A_1^* when all eigenvalues of $DT_1(A_1^*)$ have real parts less than 1.

Next consider the equation for A_0. Computing both the differential of $T_0(A_0, A_1)$ and the corresponding Jacobian $DT_0(A_0)$ analogously, we find

$$DT_0(A_0, A_1) = I \otimes \left[\Lambda_2\left(I - A_1^2\right) + \Lambda_3(I - A_1^3)\right](I - A_1)^{-1} \tag{60}$$

Provided that the motion of A_1 converges to A_1^*, the equation for A_0 is locally stable at (A_0^*, A_1^*) when the eigenvalues of $DT_0(A_0^*, A_1^*)$ have real parts less than one. This result is is in line with Proposition 5.

4 Numerical Results

4.1 Calibration

We calibrate the model at a quarterly frequency using values common in the monetary policy literature. In particular, we follow Galí (2009) and references therein and set the discount factor equal to $\beta = 0.99$. The inverse elasticity of intertemporal substitution σ is set equal to 1. On the supply side we set the price elasticity of demand between intermediate goods ϵ to 6, which corresponds to a steady state mark-up of 20%. Further, we assume that $\theta = 0.75$, which is consistent with an average period of one year between price adjustments (see Galí, 2009). The calibration of the interest rate rule follows Taylor (1993) so that we set $\phi = 1.5$, however, we assume a positive inflation target $\bar{\pi}$ of 5% annualized. Moreover, we begin with a forward guidance horizon of one year,so that $q = 4$, which is broadly in line with Hanson and Stein (2015) and Swanson and Williams (2014), who argue that the Fed's forward guidance strategy operates with a roughly two-year horizon.

For our behavioral parameters, however, there is less consensus about the calibration. Thus, we choose a forward looking horizon of the private households of $N = 4$, which corresponds to a one year horizon for which households form explicit expectation. Further, we set the intensity of choice

b equal to 2500, but provide also simulations for different values.[26] The memory parameter ρ is set equal to 0. Concerning the gain parameter ω in the adaptive learning rule, we follow Anufriev and Hommes (2012*a*,*b*) and set it equal to 0.65. This coefficient was found to predict individual forecasting behavior in learning-to-forecast experiments reasonably well.

Lastly, there is no assumed persistence in either demand or supply shock. In fact, the model produces reasonable degrees of persistence in key macro-variables without the need of highly persistent shocks (De Grauwe, 2012). Also, the distribution of the white noise processes are not assumed to be very dispersed and there is no covariance between the structural shocks.

The benchmark calibration of the main model parameters is summarized in Table 1. If the calibration differs from the benchmark case it will be made explicit. The simulations below are chosen to illustrate our *key observations*. If not stated differently, the central bank estimates a bivariate VAR(1) model.

Table 1: Parameter calibration

Parameter	Value	Description	Source
β	0.99	Discount factor	Galí (2009)
N	4	Forward-looking horizon of private households	
σ	1	Inverse elasticity of intertemporal substitution	Galí (2009)
θ	0.75	Calvo parameter	Galí (2009)
ϵ	6	Price elasticity for intermediate goods	Galí (2009)
κ	0.0572	Slope of the NKPC	Galí (2009)
		(Composite parameter for price rigidity)	
ϕ	1.5	Central bank's inflation response	Taylor (1993)
$\bar{\pi}$	0.05	Central bank's inflation target (annualized)	
$\tilde{\pi}$	0.08	Central bank's inflation tolerance (annualized)	
q	4	Central bank's forward guidance horizon	
b	2500	Intensity of choice	
ρ	0	Memory parameter performance measure	
ω	0.65	Gain parameter of adaptive learners	Anufriev and Hommes (2012*a*,*b*)
ρ_x	0	Persistence parameter in demand shock	De Grauwe (2012)
ρ_π	0	Persistence parameter in supply shock	De Grauwe (2012)
ρ_i	0	Persistence parameter in monetary policy shock	De Grauwe (2012)
σ_x	0.005	Std. deviation in demand shock	De Grauwe (2012)
σ_π	0.005	Std. deviation in supply shock	De Grauwe (2012)
σ_i	0.0005	Std. deviation in monetary policy shock	

4.2 Simulations

In this section we illustrate our findings of the numerical analysis. In fact, it turns out that our behavioral macro model allows for a rich set of dynamics, so that we restrict ourselves and only present our key observations.

The first recurrent observation is the existence of *boom* and *bust* cycles, that are created−or at least amplified−by monetary policy mistakes. In particular, if the central bank (systematically) underestimates inflation over a prolonged period of several quarters, it will pursue an interest rate

[26] Note that the calibration of the intensity of choice crucially depends on both the definition and the unit of measurement of the fitness measure. In our simulations a 1% deviation of inflation from steady state is measured as 0.01, which corresponds a squared forecast error of 0.0001. Hommes and Lustenhouwer (2015) therefore argue that an intensity of choice of 40.000 should not be considered as exceptionally large. Since we, however, define the fitness measure as the some of squared forecast errors resulting from output gap, inflation and interest rate expectations, a lower value of b results in a similar degree of sensitivity. Likewise, De Grauwe (2011) sets $b = 1$ and Anufriev and Hommes (2012*a*) sets $b = 0.4$, because both have different units of measurement.

policy that is not anti-inflationary enough. As a result, the economy overheats. Once the central bank recognizes this, nominal rates are appropriately increased, which, however, can choke off the economy if done too abruptly. Figure 3 shows such a pattern. The upper-left panel depicts the true inflation (blue) and the central bank's real-time estimate (red-dashed). Further, the lower-left panel plots both the interest rate actually set by the central bank that is consistent with its expectations (blue), as well as the nominal interest rate that would have been set if the central bank can observe current inflation, as usually assumed under rational expectations. Evident from the Figure, the central bank underestimates inflation and thus inadvertently pursues policies that are not anti-inflationary enough. Arguably, these policy mistakes can occur for two reasons: (i) the central bank finds itself in a early stage of learning (i.e. the central bank's VAR coefficients have not yet converged) and therefore forecast may be *inaccurate* or (ii) private households coordinated on adaptive expectations, which is ignored by the central bank due to *model misspecification*. In both cases the drift of inflation and/or output gap induced by the initial policy mistake leads to further coordination away from the credibility believer heuristic, which leads to self-fulfilling expectations that amplify the drift. That is, these periods with a too weak response on inflation (due to underestimation of inflation, but not due to a too small response coefficient ϕ) are characterized by a decline in central bank credibility.[27] This result is in line with recent work in progress by Lubik and Matthes (2016) who study the Fed's interest rate policy during the high-inflation period of 1970s (also known as the *Great Inflation*). Using Bayesian estimation methods the authors calibrate a stylized New Keynesian model to argue that the Great Inflation is the result of equilibrium indeterminacy, inadvertently arising from an optimizing central bank that was constraint by uncertainty about the structure of the economy and measurement error in real-time data. Both these constraints are also present in our paper, although we interpret the structural uncertainty as the central bank not being aware of the time-variation in aggregate expectations. The latter is broadly in line with Mankiw et al. (2003), who document substantial disagreement in inflation expectations (here captured by the permanent heterogeneity in private households'expectations) as well as the autocorrelated forecast error of individual households (here captured by the bounded rationality of private households).

The second observation brings us to our main research question: What are the effects of forward guidance if households are heterogeneous and boundedly rational? In particular, we find that forward guidance is a useful tool in stabilizing the economy. Figure 4 illustrates this finding by comparing two time series for our macro-variables: One with no guidance announcements in which the economy enters a deflationary spiral (*red*), and another, where we allow for forward guidance and the economy is able to recover (*blue*). The crucial point is the jump in central bank credibility (i.e. the sudden increase in the fraction of credibility believers) induced by the forward guidance policy. The reason is twofold. First, the publication of the central bank's own projections for inflation and output gap turned out to be more accurate than the private sector's own (adaptive) beliefs, which increases the bank's credibility *ex post*. This switching away from the adaptive expectations heuristic puts a halt to the diverging and self-fulfilling process in inflation and output gap. Second, the announced lower-for-long policy additionally lowers the aggregate real interest rate expectations and thus lifts the economy out of the liquidity trap. Lastly, it is important to mention that for this example shown in Figure 4 both types of guidance were needed to stabilize

[27] In longer simulations we found that these boom and bust cycles can also occur if the central bank's VAR coefficients seemed to have nearly converged to some level. In other words, the model misspecification of the central bank due to the presence of heterogeneous households (reason (ii) mentioned above) allows for policy mistakes at any point in time.

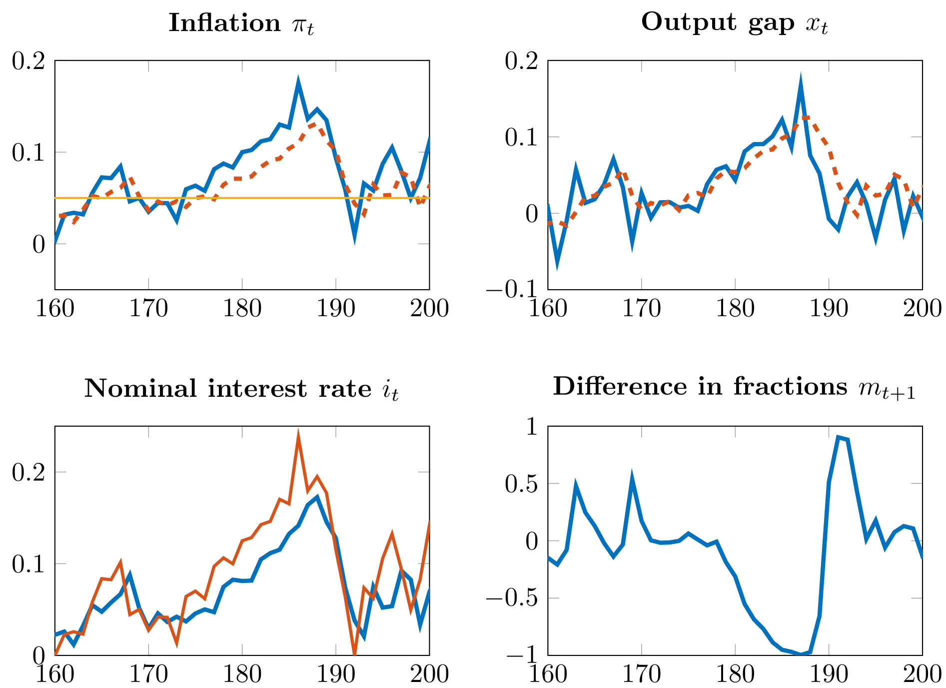

Figure 3: The central bank underestimates inflation (blue line versus red dashed line in upper left panel) and thus inadvertently pursues policies that are not anti-inflationary enough. Inflation target of 5% in orange. Central bank expectations are red-dashed. The nominal interest rate implied by *contemporaneous* information red.

the economy.

To further assess the role of both Delphic and/or Odyssean forward guidance in preventing deflationary spirals and thus stabilizing the economy, we run a total of 10,000 simulations with a length of 50 years (200 quarters) and compute the likelihood of a deflationary spiral. Additionally, a pre-period of 10 years (40 quarters) is simulated to initialize the central bank's learning algorithm.[28] A deflationary spiral is then defined as a liquidity trap with diverging inflation and/or output gap with values in excess of -100%.[29] The results are summarized in Table 2. It shows that both Delphic and Odyssean forward guidance jointly can decrease the likelihood of deflationary spirals by roughly 15 percentage points.[30] Moreover, the variation in output gap and inflation is reduced. However, Table 2 also reveals that the major contribution in lowering the likelihood of deflationary spirals de facto comes from the Delphic guidance.[31] Arguably, Delphic guidance alone is on average even more effective. Contrary, the effects of Odyssean guidance alone appear modest, decreasing the likelihood of deflationary spirals by little more than 4 percentage points.

Additionally, we ask the question how often either guidance policy—or both jointly—could

[28] In the initialization period we assume the standard contemporaneous Taylor type rule, that is the central bank perfectly observers current period endogenous variables.

[29] The results do not hinge on the exact level of the threshold.

[30] Arguably, our estimates of the likelihood of deflationary spirals appear to be rather high. We interpret our estimate to be on the upper side, yet still consider it reasonable given the length of the simulation, taking into account that we have not calibrated the model to match the moments of macroeconomic variables.

[31] We assumed that under Delphic guidance only the central bank not only publishes its inflation and output gap projections, but also the projected path of the nominal interest rate consistent with those expectations. That is, credibility believers expect the nominal interest rate for the next q periods to be given by

$$E_{1,t}i_{t+j} = \max\{0, \bar{\pi} + \phi(\pi_t^{e,cb} - \bar{\pi})\}$$

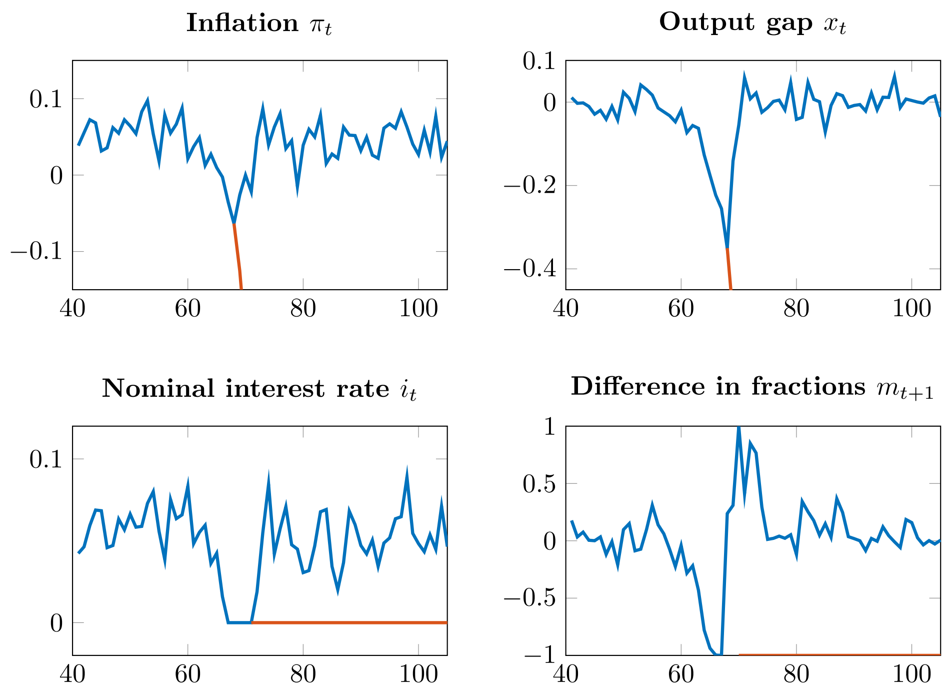

Figure 4: The economy enters the zero lower bound region in period 67 and experiences a deflationary spiral (red), while under both Delphic and Odyssean forward guidance the economy can escape the liquidity trap (blue).

Table 2: Guidance and the likelihood of deflationary spirals: Baseline calibration

	without forward guidance	Delphic guidance	Odyssean guidance	Both
Likelihood (in %)	43.15	28.34	38.96	28.86
Std. dev x_t	.054	.0431	.0513	.0452
Std. dev π_t	.0173	.0148	.0169	.0156

prevent a deflationary spiral, relative to the alternatives. In other words, we are interested in the percentage share of deflationary spirals that could be prevented under any given policy, i.e. Delphic, Odyssean or both, but not *ceteris paribus* under any other. These calculations are presented in Table 3. The table states, for instance, that in only 8.77% (40.02%) of cases both types of guidance jointly prevented a deflationary spiral, while Delphic (Odyssean) guidance alone was not sufficient. Interestingly, we find that in 10.76% of the cases, Delphic guidance alone was able to prevent a deflationary spiral, while both types of guidance were not. For this reason, we conclude that Delphic guidance is more efficient in preventing deflationary spirals as compared to Odyssean guidance alone. Moreover, we find weak evidence suggesting that Delphic guidance is even more efficient than both policies combined.

Finally, it is to mention that we also observed certain drawbacks of the guidance announcements. Specifically, we find that forward guidance announcements, particularly those of the Odyssean type, can lead to higher volatility in the periods following the announcement and even create itself boom and bust cycles. This result appears to be in sharp contrast to those found above, yet we try to argue below, that the sign of the effects (i.e. whether they are beneficial or harmful to

Table 3: The Likelihood (in %) of a deflationary spirals being prevented by the policy in baseline scenario, but *ceteris paribus* not by any of the alternative policies.

	Alternative policy		
Baseline	Both	Delphic	Odyssean
Both	0	8.77	40.02
Delphic	10.76	0	44.07
Odyssean	3.72	4.80	0

macroeconomic stability) depends on fundamentals.[32] To illustrate our point, consider Figure 5, where allowing for both Odyssean and Delphic forward guidance (*red* in Figure 5) increases the volatility in output gap and inflation ceteris paribus, by amplifying the cycles relative to the no guidance case (*blue* in Figure 5). In particular, the standard deviation for the output and inflation increases from 0.0105 and 0.0078 to 0.0154 and 0.0089, respectively. Again, we decompose the total effect into the increase arising from Delphic (*violet* in Figure 5) and Odyssean (*orange* in Figure 5) forward guidance, to better assess the source of volatility. It turns out that restricting the central bank to use only Odyssean guidance increases the volatility of output gap and inflation to 0.0205 and 0.0107, respectively, while allowing only for Delphic guidance yields 0.0113 and 0.008. Therefore, we conclude that the aforementioned increase in macroeconomic volatility is mainly induced by the central bank's *lower-for-long* policy. It is important to understand that aggregate expectations of the private sector were "optimistic enough" such that the economy is not locked into a liquidity trap, even in the absence of guidance. Hence, a commitment to lower interest rates is not needed, but rather leads to unnecessarily high levels of volatility. However, note that under only Delphic guidance inflation drops the most in period 71 and 72. In fact, this deflationary period results from what we interpret as the *communication channel* of monetary policy: the central bank itself expects deflation which it communicated to the credibility believers, which lowers private sectors' aggregate expectations relative to the scenario without guidance. Only in combination with the accommodative Odyssean guidance can this decline be limited.

To put our numerical results in a nutshell, our model predicts (i) that the combination of real-time learning, model misspecification and imperfect credibility of the central bank gives rise to policy mistakes that can result in periods of high inflation, (ii) forward guidance announcements are effective in preventing the economy from entering a deflationary spiral, notably those announcements of the Delphic type, which can decrease the probability by around 15 percentage points, however, (iii) forward guidance, particularly the lower-for-long policy, can also create additional macroeconomic volatility.

4.3 Constant gain learning

In this subsection we aim to answer the question of how important the *model misspecification* in driving our results. First, recall that this misspecification is coming from the fact that the central bank believes the aggregate law of motion to have time-invariant coefficients. In other words, the central bank believes that the economy is not subject to time switches in the sentiment of economic agents, here modeled by the time-variation in the distribution of aggregate expectations of inflation, output gap and the nominal interest rate. Next, we relax this assumption and allow

[32] With fundamentals we here mean past distributions of inflation and output gap expectations with are fundamental to the dynamics of the economy.

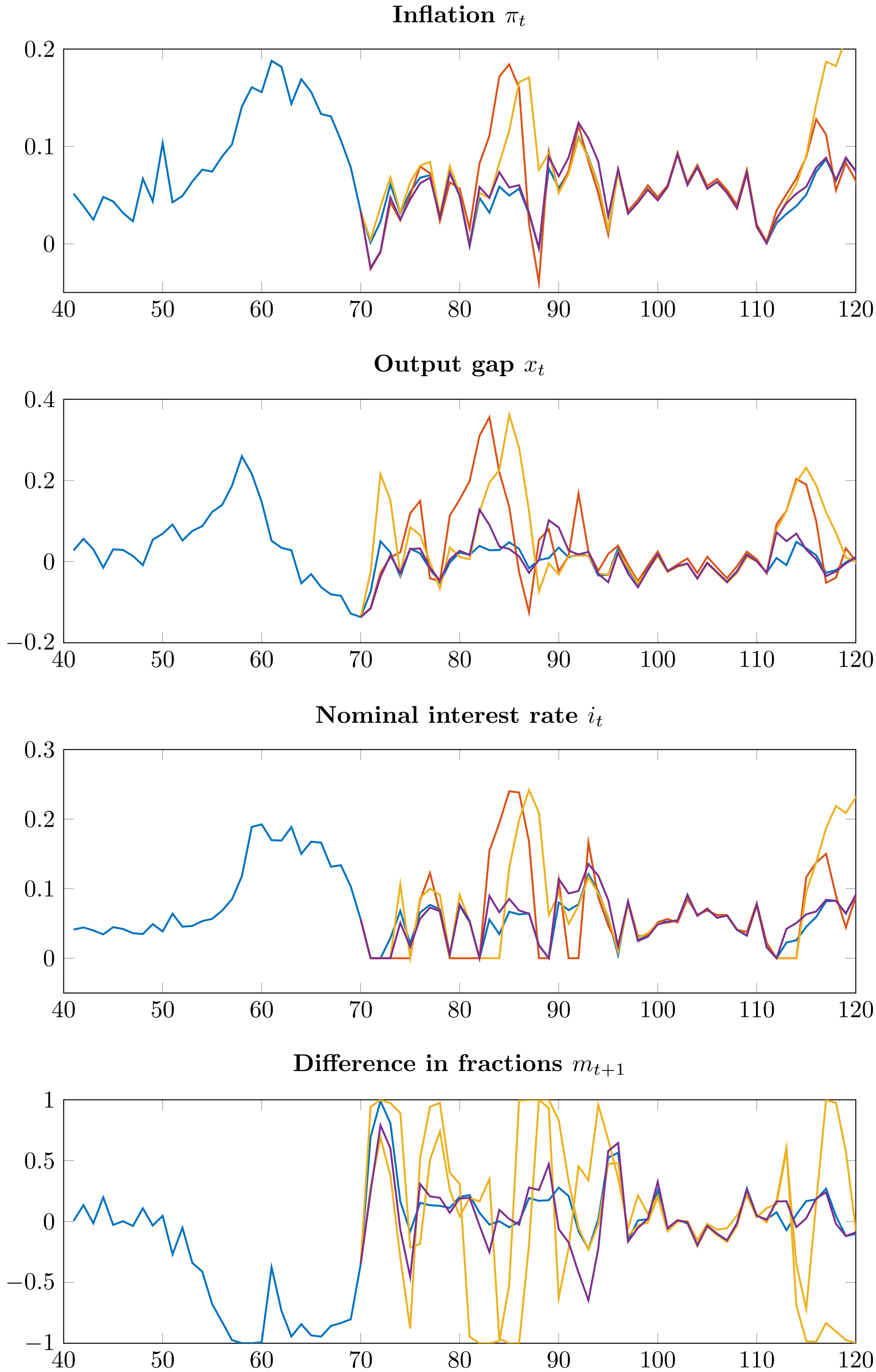

Figure 5: No guidance (blue). Delphic and Odyssean guidance (red). Odyssean guidance only (orange). Delphic guidance only (violet)

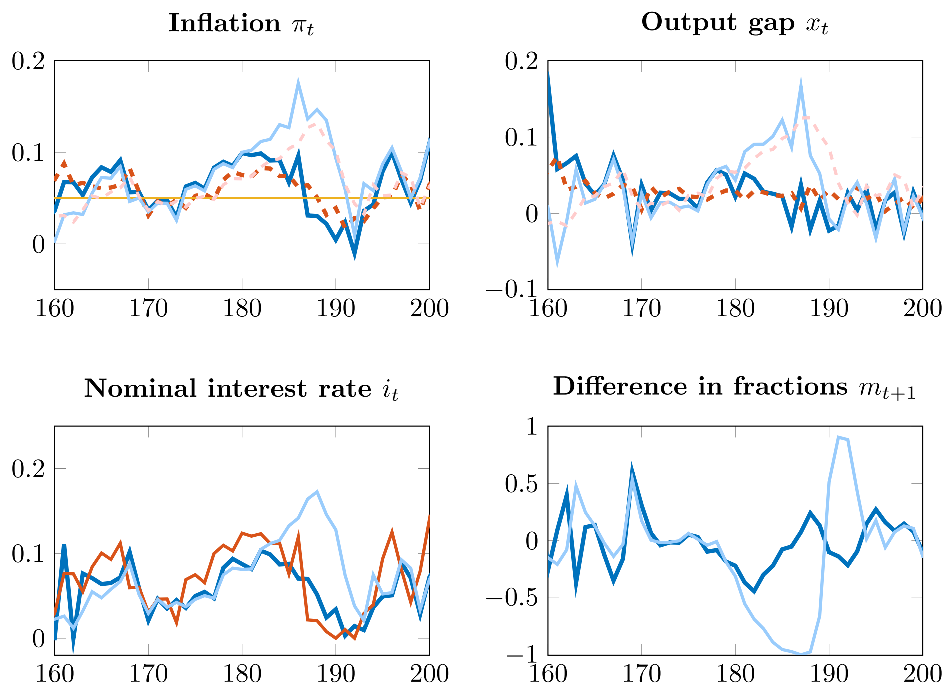

Figure 6: Counterfactual to Figure 3: Macroeconomic development under constant (strong colors) versus decreasing gain learning (lighter colors) of the central bank.

the central bank to estimate its VAR model using a *constant* gain, such that coefficients in the central bank's PLM are allowed to be time-varying.[33] In this case, the central bank is potentially able to recognize drifts in inflation earlier than under decreasing gain learning (DGL), and thus might pursue policies which are more anti-inflationary.

In what follows, we present a counterfactual analysis to the simulation presented in Figure 3. Using the same random seed and holding everything else equal, we change the gain coefficient from the decreasing gain $\gamma_t = \frac{1}{t}$ to a constant gain of $\gamma_t = \gamma$.[34] The resulting time series are presented in Figure 6. We observe that under constant gain learning (CGL), the central bank predicts slightly higher inflation around period 180, better capturing the effect the increase in adaptive learners has on the aggregate law of motion. The resulting (slightly) higher nominal interest rates are then putting a halt on inflation and output gap from further increasing, which boosts the central bank's credibility and prevents further households from switching to the adaptive learning rule. Therefore, we conclude that CGL reduces the chance of interest rate policies that are not anti-inflationary enough.

5 Robustness

In this section we are going to change some of the coefficients which are either crucial for our results, or for which no common values exist. Specifically, we will allow for different forward looking horizons N and different intensities of choice b in Section , while changing the length of the forward guidance horizon in Section 5.2.

[33] Constant gain learning is similar to a rolling window estimation, since past observations receive a geometrically declining weight.

[34] In our numerical estimations we set $\gamma = 0.02$, however, the results are robust to different choices of the constant gain parameter, which still can be considered reasonable.

5.1 Different parameterization

One of our main assumptions is that private households have a longer, but finite horizon for which they form explicit expectations as proposed by Branch et al. (2012). Generally, the stability of the dynamic system in normal times depends crucially on N. As shown in Appendix C for the special case with $\omega = 1$, an increase in N will, de facto, lead to an increase in the eigenvalues of the system. Nearly self-fulfilling expectations and slow convergence to the fundamental steady state would be the result.

Also the intensity of choice is a parameter for which no consensus among behavioral economists exists. As already discussed in the Footnote 26 the calibration of b crucially depends on both the definition and the unit of measurement of the fitness measure U_t. However, the model dynamics depend on the size of the intensity of choice in a non-trivial way. Indeed, we can think of two opposing effects resulting from an increase in the intensity of choice. Firstly, a lower intensity of choice enlarges the region of initial values for which recovery occurs, because when output gap and inflation are very low for some periods, there will still be a significant fraction of private households who believe in the central bank's ability to move the economy back into the fundamental steady state and thereby exercise an upward pressure on both inflation and output. Contrary, a lower intensity of choice can be destabilizing, if the economy is in a liquidity trap, with inflation and output gap falling. Then, even when at some point adaptive expectations perform worse, a lower intensity of choice will lead to the central bank credibility rising very slowly and thus potentially preventing the system to converge back to the fundamental steady state.

5.2 Different forward guidance horizons

Above we assumed that the forward guidance horizon of the central bank q equals the forward-looking horizon of the agents N. In this section we will show that differing forward guidance horizons has implications for the steady states, but not on the learnability results. Note, that forward guidance announcements too far in the future−those that exceed the private households forward-looking horizon−are not taken into account. Therefore, we focus on those cases in which $q < N$. Equation (46) becomes

$$y_t \equiv \begin{pmatrix} x_t \\ \pi_t \end{pmatrix} = \tilde{\Lambda}_0(m_t) + \Lambda_1(m_t)\begin{pmatrix} x_{t-1} \\ \pi_{t-1} \end{pmatrix} + \Lambda_2(m_t)\begin{pmatrix} x^{e,cb}_{t+1} \\ \pi^{e,cb}_{t+1} \end{pmatrix} + \ldots + \Lambda_{q+1}(m_t)\begin{pmatrix} x^{e,cb}_{t+q} \\ \pi^{e,cb}_{t+q} \end{pmatrix} \tag{61}$$

where matrices $\Lambda_1(m_t), \ldots, \Lambda_q(m_t)$ remain unchanged, but the vector $\tilde{\Lambda}_0(m_t)$ becomes

$$\tilde{\Lambda}_0(m_t) = \begin{pmatrix} \left[\left(\frac{1-\beta}{\kappa} + \sigma^{-1}\right)\left(\frac{1+m_t}{2}\right) - \frac{(N-1)(\phi-1)}{\sigma}\right]\bar{\pi} \\ \left[\left(1 + \beta^{q+1} - \beta + \frac{\kappa}{\sigma}\right)\left(\frac{1+m_t}{2}\right) - \frac{\kappa[(N-1)(\phi-1)]}{\sigma}\right]\bar{\pi} \end{pmatrix} \tag{62}$$

Therefore, the ALM is

$$y_t = \left[\tilde{\Lambda}_0 + \sum_{j=1}^{q} \Lambda_{j+1}\left(I - A_1^{j+1}\right)\bar{A}_0\right] + \left[\Lambda_1 + \sum_{j=1}^{q} \Lambda_{j+1}A_1^{j+1}\right] y_{t-1} \tag{63}$$

The system's stability is again determined by the matrix expression in the latter square brackets. Similarly as before, the question of the existence of a fixed point of the T-map (in particular, a fixed point for A_1) arises, . However, note that the zero lower bound steady state for $q < N$ does only exist with all agents being naive (i.e. $m^* = 0$), as in the case without forward guidance. The learnability results readily carry over from Section 3.2.

6 Conclusion

In this paper, we provide a model with endogenous central bank credibility to study the effectiveness of forward guidance under both heterogeneous and boundedly rational expectations. In particular, we model heterogeneity in expectations across households, but also with respect to the central bank. Further, private households use finite Euler equation learning and switch between simple forecasting heuristics, which households use for forecasting. The central bank, on the other hand, estimates a VAR model, which it updates recursively.

In our baseline model, the central bank uses recursive least squares with a decreasing gain, thereby assuming the economy to follow a linear model with time-invariant coefficients. More intuitively, the central bank is unaware of the time-variation in aggregate expectations which enter the economy in a nonlinear fashion. As a result of this model misspecification, the central bank learns the true model if and only if the fractions of households using a specific forecasting heuristic converges, i.e. aggregate expectations become time-invariant. Otherwise, the economy attains a restricted perception equilibrium.

The key novelty of our paper is that we make the effectiveness of forward guidance depending on the central bank's credibility, which itself is endogenously determined by how well the central bank did in achieving its target in recent past. That is, an endogenous fraction of private households incorporate the central bank's guidance announcements into their expectation formation, while the rest of the population forms expectations in an adaptive fashion. Forward guidance is then acting through two channels. First, the Delphic guidance can help to coordinate aggregate expectations if the central bank's own projections are more accurate than those of the adaptive learners. The idea is that, once the central bank publishes its (potentially more accurate) projections, households observe the forecasts of the central bank, which can ex post increase its credibility and more households coordinate on the credibility believers rule. Second, Odyssean guidance depresses expected real interest rates of the credibility believers thereby increasing demand already today. However, the effectiveness of guidance depends crucially on the central bank's credibility.

Our numerical simulations suggest that the combination of real-time learning, model misspecification and imperfect credibility of the central bank gives rise to policy mistakes that can result in periods of high inflation, potentially explaining periods such as the Great inflation in the 1980s in the United States. These policy mistakes are, however, not because of too weak (or too strong) responses to inflation, but due to the aforementioned combination of reasons. Further, allowing the central bank to use constant gain learning mitigates the problem, since it relaxes the informational constraint of the central bank. In other words, the PLM of the central bank is allowed to be time-varying, such that the central bank can better account for the time-variation in aggregate expectations. When allowing for forward guidance announcements, we find that both Delphic and Odyssean forward guidance are effective in preventing the economy from entering a deflationary spiral. Notably, announcements of the Delphic type can decrease the probability of a deflationary spiral by around 15 percentage points, while Odyssean guidance alone can only reduce the probability by roughly 4 percentage points. A drawback of the guidance policies is, however, that forward guidance, and particularly the Odyssean guidance, can under certain conditions also create additional macroeconomic volatility. Specifically, if a lower-for-long policy was not needed, doing so leads to higher volatility in inflation and output gap.

References

Adam, Klaus (2007), 'Experimental evidence on the persistence of output and inflation*', *The Economic Journal* **117**(520), 603–636.

Anufriev, Mikhail and Cars Hommes (2012*a*), 'Evolution of market heuristics', *The Knowledge Engineering Review* **27**(2), 255–271.

Anufriev, Mikhail and Cars Hommes (2012*b*), 'Evolutionary selection of individual expectations and aggregate outcomes in asset pricing experiments', *American Economic Journal: Microeconomics* **4**(4), 35–64.

Aoki, Kosuke and Kalin Nikolov (2006), Rule-based monetary policy under central bank learning, *in* 'NBER International Seminar on Macroeconomics 2004', The MIT Press, pp. 145–195.

Assenza, Tiziana, Peter Heemeijer, Cars H Hommes and Domenico Massaro (2013), 'Individual expectations and aggregate macro behavior'.

Blinder, Alan S (1999), Central bank credibility: why do we care? how do we build it?, Technical report, National Bureau of Economic Research.

Branch, William A (2004), 'The theory of rationally heterogeneous expectations: Evidence from survey data on inflation expectations', *The Economic Journal* **114**(497), 592–621.

Branch, William A (2007), 'Sticky information and model uncertainty in survey data on inflation expectations', *Journal of Economic Dynamics and Control* **31**(1), 245–276.

Branch, William, George W Evans and Bruce McGough (2012), 'Finite horizon learning', *CDMA, Working Paper Series CDMA12/04* .

Brock, William A and Cars H Hommes (1997), 'A rational route to randomness', *Econometrica: Journal of the Econometric Society* **65**, 1059–1095.

Campbell, Jeffrey R, Charles L Evans, Jonas DM Fisher, Alejandro Justiniano, Charles W Calomiris and Michael Woodford (2012), 'Macroeconomic effects of federal reserve forward guidance [with comments and discussion]', *Brookings Papers on Economic Activity* pp. 1–80.

Carlstrom, Charles T, Timothy S Fuerst and Matthias Paustian (2012), 'Inflation and output in new keynesian models with a transient interest rate peg'.

Carroll, Christopher D (2003), 'Macroeconomic expectations of households and professional forecasters', *the Quarterly Journal of economics* pp. 269–298.

Cochrane, John H (2015), The new-keynesian liquidity trap, Technical report.

Cole, S (2015), Learning and the effectiveness of central bank forward guidance, Technical report, Working paper, UC-Irvine.

De Grauwe, Paul (2011), 'Animal spirits and monetary policy', *Economic Theory* **47**(2-3), 423–457.

De Grauwe, Paul (2012), *Lectures on Behavioral Macroeconomics*, Princeton University Press.

Del Negro, Marco, Marc P Giannoni and Christina Patterson (2012), 'The forward guidance puzzle', *FRB of New York Staff Report* (574).

Evans, George W, Eran Guse and Seppo Honkapohja (2008), 'Liquidity traps, learning and stagnation', *European Economic Review* **52**(8), 1438–1463.

Evans, George W and Seppo Honkapohja (2001), *Learning and expectations in macroeconomics*, Princeton University Press.

Evans, George W and Seppo Honkapohja (2006), 'Monetary policy, expectations and commitment*', *The Scandinavian Journal of Economics* **108**(1), 15–38.

Ferrero, Giuseppe and Alessandro Secchi (2009), 'The announcement of monetary policy intentions', *Bank of Italy Temi di Discussione (Working Paper) No* **720**.

Ferrero, Giuseppe and Alessandro Secchi (2011), 'Central bank's macroeconomic projections and learning', *SSRN Working Paper Series* .

Galí, Jordi (2009), *Monetary Policy, inflation, and the Business Cycle: An introduction to the new Keynesian Framework*, Princeton University Press.

García-Schmidt, Mariana and Michael Woodford (2015), Are low interest rates deflationary? a paradox of perfect-foresight analysis, Technical report, National Bureau of Economic Research.

Haberis, Alex, Richard Harrison and Matt Waldron (2014), Transitory interest-rate pegs under imperfect credibility, Technical report.

Hanson, Samuel G and Jeremy C Stein (2015), 'Monetary policy and long-term real rates', *Journal of Financial Economics* **115**(3), 429–448.

Hommes, Cars (2011), 'The heterogeneous expectations hypothesis: Some evidence from the lab', *Journal of Economic dynamics and control* **35**(1), 1–24.

Hommes, Cars and Gerhard Sorger (1998), 'Consistent expectations equilibria', *Macroeconomic Dynamics* **2**(03), 287–321.

Hommes, Cars H and Joep Lustenhouwer (2015), Inflation targeting and the zero lower bound under endogenous credibility, Technical report.

Hommes, Cars, Joep Sonnemans, Jan Tuinstra and Henk Van de Velden (2005), 'Coordination of expectations in asset pricing experiments', *Review of Financial studies* **18**(3), 955–980.

Honkapohja, Seppo and Kaushik Mitra (2005), 'Performance of monetary policy with internal central bank forecasting', *Journal of Economic Dynamics and Control* **29**(4), 627–658.

Honkapohja, Seppo and Kaushik Mitra (2006), 'Learning Stability in Economies with Heterogeneous Agents', *Review of Economic Dynamics* **9**(2), 284–309.

Honkapohja, Seppo and Kaushik Mitra (2015), 'Comparing inflation and price level targeting: the role of forward guidance and transparency'.

Honkapohja, Seppo, Kaushik Mitra and George W Evans (2012), 'Notes on agents' behavioral rules under adaptive learning and studies of monetary policy'.

King, Robert G (2009), Discussion of "limitations on the effectiveness of forward guidance at the zero lower bound", Technical report.

Lubik, Thomas and Christian Matthes (2016), Indeterminacy and learning: An analysis of monetary policy in the great inflation, Technical report.

Mankiw, N Gregory, Ricardo Reis and Justin Wolfers (2003), Disagreement about inflation expectations, Technical report, National Bureau of Economic Research.

Marcet, Albert and Thomas J Sargent (1989), 'Convergence of least squares learning mechanisms in self-referential linear stochastic models', *Journal of Economic theory* **48**(2), 337–368.

Orphanides, Athanasios (2001), 'Monetary policy rules based on real-time data', *American Economic Review* pp. 964–985.

Pfajfar, Damjan (2009), Heterogeneous expectations in macroeconomics, PhD thesis, University of Cambridge.

Pfajfar, Damjan and Blaz Zakelj (2011), Inflation expectations and monetary policy design: Evidence from the laboratory, Technical report, CentER Discussion Paper Series.

Pfajfar, Damjan and Emiliano Santoro (2010), 'Heterogeneity, learning and information stickiness in inflation expectations', *Journal of Economic Behavior & Organization* **75**(3), 426–444.

Sargent, Thomas J (1999), *The conquest of American inflation*, Princeton University Press.

Simon, Herbert A (1984), 'On the behavioral and rational foundations of economic dynamics', *Journal of Economic Behavior & Organization* **5**(1), 35–55.

Swanson, Eric T and John C Williams (2014), 'Measuring the effect of the zero lower bound on medium-and longer-term interest rates', *American Economic Review* **104**(10).

Taylor, John B (1993), Discretion versus policy rules in practice, *in* 'Carnegie-Rochester conference series on public policy', Vol. 39, Elsevier, pp. 195–214.

A Rational expectations solution

In the absence of autocorrelated shocks the model under rational expectations of both private agents and the central bank coincides with the model under perfect foresight (at least under determinacy). Therefore, we substitute $E_t z_{t+j} = z_t = z^*$ for $z \in \{x, \pi, i\}$ and $j = 1, \ldots N$ into the model described by Equations (3) and (4) to find

$$\begin{cases} x^* &= x^* - \frac{N}{\sigma}\big(i^* - \pi^*\big) \\ \pi^* &= \beta^N \pi^* + \frac{1-\beta^N}{1-\beta}\kappa x^* \\ i^* &= \bar{\pi} + \phi(\pi^* - \bar{\pi}) \end{cases} \quad \Leftrightarrow \quad \begin{cases} x^* &= \frac{1-\beta}{\kappa}\bar{\pi} \\ \pi^* &= \bar{\pi} \\ i^* &= \bar{\pi} \end{cases} \tag{A.1}$$

That is, the rational expectations equilibrium is characterized by inflation as well as nominal interest rates equal to the inflation target, the latter is because we normalized equilibrium real interest rates to zero, and output gap being (slightly) positive because $\frac{1-\beta}{\kappa}$ is usually calibrated to be small, but positive.

B Stability of the central bank's VAR model

Stability of the bivariate VAR(1) requires that the roots of $(1 - a_{11}L)(1 - a_{22}L) - (a_{12}a_{21}L^2)$ lie outside the unit circle. As stated in the proof to Proposition 1, this condition is also a necessary and sufficient condition for local stability of the fundamental steady state. To find parameter conditions for which the fundamental steady state is locally stable define the characteristic equation

$$H(L) = (1 - a_{11}L)(1 - a_{22}L) - (a_{12}a_{21}L^2) \tag{B.1}$$

For $L = 0$ we have that $H(0) = 1 > 0$. Thus, the parameters must satisfy $H(1) > 0$ as well as $H(-1) > 0$ to have that both eigenvalues lie outside the unit circle. Note that the difference in fractions m_t is bounded, taking on values on the interval $[-1, 1]$. Further, with an infinite intensity of choice, i.e. $b = \infty$, m_t is restricted to either of the values -1, 0 or 1. Analyzing local stability properties conditionally on these three cases allows us to reduce the five-dimensional state vector to $[x_t, \pi_t]'$. Thus, let us for now confine ourselves to the analysis with an infinite intensity of choice. We consider the three fixed point $\mathbf{a}^*$ of the T-map (26) that satisfies $T(\mathbf{a}^*) = \mathbf{a}^*$ for the difference in fractions $m \in \{-1, 0, 1\}$. If $m = -1$, i.e. all agents are naive, the fixed point is given by

$$a_{10} = \frac{N(\phi - 1)}{\sigma}\bar{\pi} - \frac{\phi a_{20}}{\sigma} \tag{B.2}$$

$$a_{11} = 1 - \frac{\phi a_{21}}{\sigma} \tag{B.3}$$

$$a_{12} = -\frac{(N-1)\phi - N}{\sigma} - \frac{\phi a_{22}}{\sigma} \tag{B.4}$$

$$a_{20} = \frac{\kappa N(\phi - 1)\bar{\pi}}{\sigma + \kappa\phi} \tag{B.5}$$

$$a_{21} = \frac{\kappa\sigma}{\sigma + \kappa\phi}\left(\frac{1 - \beta^N}{1 - \beta}\right) \tag{B.6}$$

$$a_{22} = \frac{\sigma}{\sigma + \kappa\phi}\left(\beta^N - \frac{\kappa\big[(N-1)\phi - N\big]}{\sigma}\right) \tag{B.7}$$

Given these parameters the VAR(1) has one stable eigenvector if $\phi > 1$. The other eigenvalue is stable only if

$$\phi > \phi_{min} \equiv \frac{\kappa N(2\beta - \beta^N) - 2(1-\beta^N)\sigma}{\kappa N(2\beta - 1 - \beta^N) + 4\kappa(1-\beta)}$$

Using the calibration chosen in our simulations, for instance, ϕ_{min} is roughly 23, which lies outside any reasonable parameter space. Contrary, if $m = 1$, i.e. all agents are credibility believers, we have

$$a_{10} = \frac{1-\beta}{\kappa}\bar{\pi}, \quad a_{11} = 0, \quad a_{12} = 0 \tag{B.8}$$

$$a_{20} = \bar{\pi}, \quad a_{21} = 0, \quad a_{22} = 0 \tag{B.9}$$

That is, the VAR estimates reduce to point estimates of the fundamental steady state. Note, however, that $m = 1$ cannot be an equilibrium, as naive agents will always forecast correctly in equilibrium. As a result, in steady state is must be that $m \leq 0$ given an infinite (actually any positive) intensity of choice. Lastly, we consider the case when fractions are equal and $m = 0$. The fixed point is then given by

$$a_{10} = \left(\frac{1-\beta}{2\kappa} + \frac{(N+1)\phi - N}{2\sigma}\right)\bar{\pi} - \frac{\phi a_{20}}{\sigma} \tag{B.10}$$

$$a_{11} = \frac{1}{2} - \frac{\phi a_{21}}{\sigma} \tag{B.11}$$

$$a_{12} = -\frac{(N-1)\phi - N}{2\sigma} - \frac{\phi a_{22}}{\sigma} \tag{B.12}$$

$$a_{20} = \frac{\sigma}{2(\sigma + \kappa\phi)}\left(1 + \frac{\kappa\big[(N+1)\phi - N\big]}{\sigma}\right)\bar{\pi} \tag{B.13}$$

$$a_{21} = \frac{\kappa\sigma}{2(\sigma + \kappa\phi)}\left(\frac{1-\beta^N}{1-\beta}\right) \tag{B.14}$$

$$a_{22} = \frac{\sigma}{2(\sigma + \kappa\phi)}\left(\beta^N - \frac{\kappa\big[(N-1)\phi - N\big]}{\sigma}\right) \tag{B.15}$$

Using these parameters, we find the following conditions on ϕ to have that the VAR(1) is stable

$$\phi > \frac{N}{N-3} - \frac{3\sigma(1-\beta)(2+\beta^N)}{\kappa(N-3)(3\beta - 2 - \beta^N)} \quad \text{if} \quad N > 3 \tag{B.16}$$

$$\phi < -\frac{N}{3-N} + \frac{3\sigma(1-\beta)(2+\beta^N)}{\kappa(3-N)(3\beta - 2 - \beta^N)} \quad \text{if} \quad N < 3 \tag{B.17}$$

and

$$\phi > \frac{N}{N+1} - \frac{\sigma(1-\beta)(2-\beta^N)}{\kappa(N+1)(2-\beta-\beta^N)} \tag{B.18}$$

In case N is exactly equal to 3, the sign of $H(-1)$ does not depend on ϕ. If either Equation (B.16) or (B.17) hold with equality, the eigenvalue is exactly equal to -1, while it is precisely 1 if Equation (B.18) holds with equality.

C Eigenvalues of the dynamic system in *normal* times

In this section we derive the eigenvalues of the system in normal times, described by Equation (25), as a function of the steady state level of the difference in fractions m^*. The corresponding reduced-form Jacobian matrix evaluated in the *fundamental* steady state (i.e. with $\pi^* = i^* = \bar{\pi}$ and $x^* = \frac{1-\beta}{\kappa}$) but with general fractions (instead of $m^* = 0$) is given by

$$J = \begin{pmatrix} \frac{1-m^*}{2} - \frac{\phi a_{21}}{\sigma} & -\frac{(N-1)\phi - N}{\sigma}\left(\frac{1-m^*}{2}\right) - \frac{\phi a_{22}}{\sigma} \\ \kappa\left(\frac{1-\beta^N}{1-\beta}\right)\frac{1-m^*}{2} - \frac{\kappa\phi a_{21}}{\sigma} & \left(\beta^N - \frac{\kappa[(N-1)\phi - N]}{\sigma}\right)\frac{1-m^*}{2} - \frac{\kappa\phi a_{22}}{2} \end{pmatrix}$$

using Equations (31) and (32) we find the characteristic equation

$$\lambda^2 - \frac{1-m^*}{2}(1+\mu_1-\mu_2)\lambda + \left(\frac{1-m^*}{2}\right)^2[\mu_1+\mu_3] = 0$$

where

$$\begin{aligned} \mu_1 =& \frac{\sigma}{\sigma+\kappa\phi}\left(\beta^N - \frac{\kappa[(N-1)\phi - N]}{\sigma}\right) \\ \mu_2 =& \frac{\kappa\phi}{\sigma+\kappa\phi}\left(\frac{1-\beta^N}{1-\beta}\right) \\ \mu_3 =& \left(\beta^N - \frac{\kappa[(N-1)\phi - N]}{\sigma}\right)\left[1 + (\frac{\kappa\phi}{\sigma+\kappa\phi})^2\right] + \frac{\sigma}{\sigma+\kappa\phi}\left(\frac{1-\beta^N}{1-\beta}\right)\frac{\kappa[(N-1)\phi - N]}{\sigma} \end{aligned}$$

Then, we eigenvalues as a function of the difference in fractions m_t immediately follow

$$\lambda_{1/2}(m^*) = \frac{1}{2}(\frac{1-m^*}{2})\left[(1+\mu_1-\mu_2) \pm \sqrt{\left(1-\mu_1-\mu_2\right)^2 - 4\mu_3}\right] \tag{C.1}$$

Clearly, if all agents are credibility believers ($m^* = 1$) both eigenvalues are zero. However, for an increasing fraction of naive agents (decreasing m^*) the eigenvalues will increase in absolute value and potentially crossing the unit circle. In the fundamental steady state we have $m^* = 0$ (i.e. both heuristics perform equally well), so that−for N not too large−both eigenvalues remain inside the unit circle.

D Proof of Proposition 3

Proof. We begin with assuming constant fractions, i.e. $m_t = m$. with $m \in [-1, 1)$. Under this assumption, together with condition (40) (which implies that $\iota = 0$ in steady state) the dynamic system in Equations (42) and (43) evaluated in steady state reduces to

$$\begin{aligned} x^* &= \left(\frac{1-\beta}{\kappa} + \frac{1}{\sigma}\right)\bar{\pi} + \frac{N}{\sigma}\left(\frac{1-m}{1+m}\right)\pi^* \\ \pi^* &= \frac{1+m}{2}\left(1+\frac{\kappa}{\sigma}\right)\bar{\pi} + \kappa\left(\frac{1-\beta^N}{1-\beta}\right)\frac{1-m}{2}x^* + \frac{1-m}{2}\left(\beta^N + \frac{\kappa N}{\sigma}\right)\pi^* \end{aligned}$$

from which we can derive the following steady state expression for inflation π_m^* given constant fractions m

$$\pi_m^* = \frac{(1+m)^2\big(1+\frac{\kappa}{\sigma}\big)+(1-\beta^N)\big(1+\frac{\kappa}{\sigma(1-\beta)}\big)(1-m^2)}{2(1+m)-\frac{\kappa N}{\sigma}\big(\frac{1-\beta^N}{1-\beta}\big)(1-m)^2-\big(\beta^N+\frac{\kappa N}{\sigma}\big)(1-m^2)}\bar{\pi}, \quad \text{for } m \in [-1,1) \tag{D.1}$$

which has a vertical asymptote at

$$\tilde{m}_1 = \frac{-\alpha_2+\sqrt{\alpha_2^2-4\alpha_1\alpha_3}}{2\alpha_1}$$

where

$$\alpha_1 = \beta^N - \frac{\kappa N}{\sigma}\Big(\frac{\beta-\beta^N}{1-\beta}\Big) \gtrless 0$$
$$\alpha_2 = 2\Big[1+\frac{\kappa N}{\sigma}\Big(\frac{1-\beta^N}{1-\beta}\Big)\Big] > 0$$
$$\alpha_3 = 2-\beta^N - \frac{\kappa}{\sigma}\Big(\frac{2-\beta-\beta^N}{1-\beta}\Big) \gtrless 0$$

and $\tilde{m}_2 \notin [-1,1]$. Next, recognize that the numerator of (D.1) is always positive. The denominator, however, is negative for $m \in [-1,\tilde{m}_1)$ and approaches $-\infty$ for $m \to \tilde{m}_1^-$, while it is positive for $m \in (\tilde{m}_1, 1)$ and approaches ∞ for $m \to \tilde{m}_1^+$. The latter is in contradiction with condition (40), since π_m^* needs to be smaller than zero for the ZLB steady state to exist, given any positive fraction of credibility believers. Hence, we conclude that a steady state $\pi_m^* \in [0,-\infty)$ in the ZLB region exists for $m \in [-1,\tilde{m}_1)$.

Lastly, consider the case of time-varying fractions. Given a positive intensity of choice, the only possible steady state in the ZLB region features $m=-1$, since credibility believers would make systematic forecast errors in this steady state and thus all households must necessarily be naive (or adaptive learners). From Equations (3) and (4) it is easy to see that $x=\pi=i=0$ is the unique solution with corresponding expectations satisfying $Ex=E\pi=Ei=0$. However, for the this candidate steady state to be consistent with the ZLB region, we must additionally have that the ZLB constraint is binding, i.e. Equation (40) must hold. Given $x=\pi=0$ in steady state the constant $a_{20} \to 0$ such that also $\pi^{e,cb}=0$ and condition (40) reduces to $0<(1-\phi^{-1})\bar{\pi}$ which is equivalent to $\phi>1$ given $\bar{\pi}>0$. □

E Proof of Proposition 4

Proof. As in the proof to Proposition 3, we begin with assuming fixed fractions $m \in [-1,1]$. Under this assumption, the dynamic system in Equations (44) and (45) reduces to $\pi_m^* = x_m^* = 0$, once we substitute $z_t = z_{t-1} = z_t^{e,cb} = z$ for $z \in \{x,\pi\}$. That is, for $N \leq q$ the ZLB steady state exists independent of the share of credibility believers.

Under time-varying fractions, the steady state features $m=0$. The reason is that through the forward guidance announcements even the credibility believers predict correctly in in the ZLB steady state. As before, for Condition (40) to be satisfied, we must have that $\phi>1$ given $\bar{\pi}>0$. □