Inflation Targeting and Liquidity Traps under Endogenous Credibility*

Cars Hommes[†a,b] and Joep Lustenhouwer[a,b]

[a] *CeNDEF, University of Amsterdam*

[b] *Tinbergen Institute, Amsterdam*

November 2, 2015

Abstract

We derive policy implications for an inflation targeting central bank, who's credibility is endogenous and depends on its past ability to achieve its targets. We do this in a New Keynesian framework with heterogeneous agents and boundedly rational expectations. Our assumptions about expectation formations are more in line with expectations observed in survey data and laboratory experiments than the fairly restrictive rational expectations hypothesis. We find that the region of allowed policy parameters is strictly larger under heterogeneous expectations than under rational expectations. Furthermore, with theoretically optimal monetary policy, global stability of the fundamental steady state can be achieved, implying that the system always converges to the targets of the central bank. This result however no longer holds when the zero lower bound (ZLB) on the nominal interest rate is accounted for. Self-fulfilling deflationary spirals can then occur, even under optimal policy. The occurrence of these liquidity traps crucially depends on the credibility of the central bank. Deflationary spirals can be prevented with a high inflation target, aggressive monetary easing, or a more aggressive response to inflation.

JEL Classification: E52, E32, C62

Keywords: Monetary Policy, Liquidity Traps, Heterogeneous Expectations, Bounded Rationality

*The authors gratefully acknowledge the financial support of NWO (Dutch Science Foundation) Project No. 40614011 "Monetary and Fiscal Policy under Bounded Rationality and Heterogeneous Expectations", and the financial support from the EU FP7 project "Integrated Macro-Financial Modeling for Robust Policy Design" (MACFINROBODS, grant 612796).

[†]Email addresses: C.H.Hommes@uva.nl, J.E.Lustenhouwer@uva.nl

1 Introduction

Many central banks (CB's) have recently adopted some form of inflation targeting. Some CB's do this by claiming to set the interest rate such that, if the interest rate were to be kept constant, the inflation in some specified future period is expected to equal the target value. Other CB's[1] include predictions about the future path of the interest rate in their expectations. Both forms of inflation targeting can be described as "inflation forecast targeting".

Besides setting the interest rate optimally, an important aspects of inflation targeting is *managing expectations*. This is for example stressed by Woodford (2004). For the inflation targeting to be effective it is important that the the CB has enough *credibility*. If the private sector does not believe the CB when it announces an inflation target, the realized value of inflation will likely not be equal to this target. Whether the CB is likely to be believed furthermore will typically depend on whether it was able to achieve its targets in the past.

Inflation targeting is usually modeled in a New Keynesian setting under the assumption that agents have fully rational expectations. Under this assumption all agents form the same perfectly model consistent expectations, which, in the absence of shocks, implies perfect foresight. When rational expectations are assumed, there is no longer a clear role for the credibility of an inflation and output gap target inside the model. Either expectations about inflation and output coincide with the targets of the central bank and the CB has full credibility[2], or expectations are not in line with the targets and the announcements of the CB are not credible.

Rational expectations are furthermore an unrealistically strong assumption when inflation and output forecasts by price setters (i.e. the private sector) are concerned. Both surveys of consumers and professional forecasters and laboratory experiments with human subjects show that there is considerable heterogeneity in inflation forecasts (e.g. Mankiw et al., 2003, and Pfajfar and Zakelj, 2011). Assenza et al. (2014) furthermore show that in their laboratory experiments, expectations of subjects can quite accurately (both qualitatively and quantitatively) be described as switching between simple heterogeneous forecasting heuristics based on their relative past performance.

[1] E.g. the Central Bank of New Zealand

[2] Note that we refer to the credibility of the CB's targets, and not to the credibility of its future policy actions. See Section 3 for details.

In this paper we investigate monetary policy in a standard New Keynesian model (in line with Woodford, 2003 and Galí, 2002) where the assumptions of homogeneous and rational expectations are relaxed. Instead, a heuristic switching model with heterogeneous expectations is used, that allows for *endogenous credibility*. Heuristic switching models were introduced by Brock and Hommes (1997), and have since successfully been used to model heterogeneous expectations in finance and macroeconomics (Hommes, 2013). In our model agents switch between two intuitive forecasting heuristics based on *relative performance*. Branch and McGough (2010) and Cornea et al. (2013) use combinations of forecasting heuristics that are similar to ours. Other works with heuristic switching models in a macroeconomic setting include De Grauwe (2011) and Anufriev et al. (2013).

The most important forecasting heuristic can be described as "Trust the central bank". Followers of this heuristic are called fundamentalists, and expect future inflation and output gap to be equal to the targets of the central bank. The fraction of fundamentalists can be interpreted as the *credibility* of the central bank. In contrast with rational expectations models, our model therefore involves *endogenous credibility*. We assume that these fundamentalists compete with naive expectations, which uses the last observation as a best guess for future realizations of inflation and output. The naive heuristic coincides with rational expectations when inflation or output follows a random walk. If inflation or output follows a near unit root process, the naive forecast is nearly rational. Naive agents furthermore adds persistence in inflation and output gap to our model in a very simple and intuitive manner, without the need to assume heavily serially correlated shocks. Similar to the naive heurisitic, Milani (2007) has stressed that homogeneous adaptive learning also generates high persistence in inflation and output dynamics, especially under constant gain learning (Evans and Honkapohja, 2009).

Cornea et al. (2013) estimate a New Keynesian Phillips curve assuming expectations are formed by a heuristic switching model with fundamentalists and naive agents. Fundamentalists here make use of the forward looking relation between inflation and marginal cost and use a VAR approach to make inflation forecasts. Cornea et al. (2013) find that their model fits the data quite nicely and that the endogenous mechanism of switching between the two heuristics based on past performance

is supported by the data. Branch (2004, 2007) fit a heuristic switching model with amongst others a naive heuristic, and a fundamentalistic VAR heuristic to data from Michigan Survey of Consumer Attitudes an Behavior. Both these papers find clear evidence of switching between heuristics based on past performance. Branch (2004) furthermore finds that both our heuristics are present in the survey data, and Branch (2007) finds that the heuristic switching model better fits the survey data than a static sticky information model. A case study of the Volcker disinflation by Mankiw et al. (2003) furthermore nicely illustrates the presence of our two heuristics in survey data. In Figure 12 (Mankiw et al., 2003, p. 46) the evolution of inflation expectations as measured by the Michigan Survey from 1979 up to and including 1982 is plotted. They show that at the start of 1979 expectations were centered around a high inflation value. Over the next eight quarters (during which Paul Volcker was appointed chairmen of the Board of Governors of the Federal Reserve Board) the distribution of expectations clearly becomes bimodal, with a fraction of agents still expecting the same high values of inflation and another fraction expecting lower inflation. In terms of our model we can interpreted this as follows. Before Volcker was appointed the FED had very little credibility and most agents expected inflation to remain at the high values that it had been in the recent past (they used the naive heuristic). In the following quarters the FED gained more credibility and an increasing fraction of agents started to believe that Volcker would be able to drive down inflation towards its target level (more agents started to follow the fundamentalistic heuristic). Furthermore, when in 1982 actual inflation started to decline, the mass on high inflation expectations slowly started to move towards lower inflation. We can interpreted this as backward looking, naive agents believing that lower observed inflation would also mean lower inflation in the future.

Monetary policy is often modeled in the literature with a Taylor type interest rate rule. With such an interest rate rule, the CB adjusts the interest rate in response to inflation and output gap in order to steer inflation towards a long term target. There is however little consideration for the optimal paths of inflation and output gap, and the time at which the long term target should be reached. Inflation forecast targeting is also modeled in the literature, but this is a form of strict inflation targeting, where no output considerations are allowed. While CB's claim to set the

interest rate to target inflation in some specified future period, in practice they will also take the consequences on output into account. A central bank would furthermore not only like inflation to be equal to its target in the specified period, but rather in every period.

For this reason we use an interest rate rule that is derived from a loss function consisting of the expectations of all future deviations of inflation from its target and all future output gaps. By choosing its policy rate to minimize this loss function the central bank can optimize the paths of future inflation and output gap. It turns out that, in our model of boundedly rational expectations, this optimal path can be achieved with an expectation based Taylor rule, where the interest rate depends on expectations of inflation and output gap instead of contemporaneous values. Such a rule is, amongst others, used by Bullard and Mitra (2002), and the optimal policy benchmark of a similar rule is derived by Evans and Honkapohja (2003).

In the derivation of this optimal rule, no restrictions are placed on the values that can be taken by the nominal interest rate. However, in practice this instrument will never be set negative. While ignoring this zero lower bound on the interest rate may not lead to problems when analyzing monetary policy in normal times, the recent financial crisis has shown the importance of the restrictions placed by this lower bound. Due to these restrictions the CB may not be able to stimulate the economy enough after a negative shock. This may then lead to a liquidity trap, as experienced by Japan since the 1990s.

When a central bank is constrained by the zero lower bound on the nominal interest rate, it can no longer use its main instrument to conduct monetary policy, but must instead rely on forward guidance and open-mouth operations. This implies that here it is even more crucial to realistically model expectations than during economically healthy times. Our model of endogenous credibility can provide new insights in liquidity traps that could not have been obtained under rational expectations.

Closely related to our investigation of liquidity traps under bounded rationality is a series of papers by Evans et al. (2005, 2008, 2014). These authors study monetary and fiscal policy under adaptive learning in various macroeconomic models, ranging from a simple endowment economy (Evans and Honkapohja, 2005) to a more elaborate New Keynesian framework (Evans et al., 2008;

Benhabib et al., 2014). They show the existence of multiple equilibria: the target equilibrium, and an equilibrium with low inflation. The existence of this second equilibrium when a zero lower bound is introduced to the model has been initially highlighted in Benhabib et al. (2001*a*,*b*). Evans et al. (2008) furthermore show that in their model a liquidity trap arises in the form of a deflationary spiral with ever decreasing inflation and output gap. A drawback of these models is their focus on a representative agent with adaptive learning. In our model, we extend the analysis to allow for heterogeneity in expectations and endogenous credibility.

We first analyze our heuristic switching model *without* the zero lower bound on the interest rate. The main research question here is what policy parameters lead to desirable dynamics when expectations are *heterogeneous* and *boundedly rational*. It is shown that the region of policy parameters that leads to a *locally stable* fundamental steady state (with zero output gap and inflation equal to its target) is strictly larger than the region of policy parameters that gives a locally determinate rational expectations equilibrium. Even when the Taylor principle is not satisfied there could very well be convergence to the fundamental steady state in our model. Furthermore, under the set of policy parameters that minimize the central banks loss function, the fundamental steady state is unique and globally stable for any calibration of model parameters. Without the zero lower bound, the policy that minimizes the loss function can therefore indeed be considered optimal under heterogeneous expectations.

Next, we introduce the zero lower bound (ZLB) on the nominal interest rate to the above heterogeneous expectations framework, and investigate its effect on inflation and output gap dynamics. It turns out that with the zero lower bound, expectation driven liquidity traps can arise. In rational expectations models shocks to the fundamentals of the economy can lead to a temporary liquidity trap. However, as soon as the sequence of bad shocks is over, the liquidity trap is immediately resolved. In our model a one period shock to economic fundamentals can lead to a prolonged liquidity trap due to a loss in credibility of the central bank and low, self-fulfilling expectations. Mertens and Ravn (2014) highlight the distinction between expectation driven liquidity traps and fundamental liquidity traps. Depending on the magnitude of the shock and the loss in credibility, our expectation driven liquidity traps can be temporary or take the form of a

deflationary spiral with ever decreasing inflation and output gap. Deflationary spirals have recently been observed in laboratory experiments by Assenza et al. (2014), Hommes et al. (2015), and Arifovic and Petersen (2015). We show that even under optimal monetary policy deflationary spirals can occur. When the zero lower bound is accounted for, the fundamental steady state can therefore no longer be *globally* stable, but only *locally* stable, coexisting with a deflationary spiral region.

Finally, we conduct simulations to investigate policies to prevent or recover from liquidity traps. We show three deviations of theoretically optimal monetary policy that are successful in preventing liquidity traps and deflationary spirals. First of all the central bank can prevent deflationary spirals by letting the interest rate respond more aggressively to inflation than specified by optimal policy. Alternatively, the central bank can make liquidity traps less likely by increasing the inflation target and conducting aggressive monetary easing as soon as a liquidity trap is imminent.

The paper is organized as follows. In Section 2 the New Keynesian model and interest rate rule are presented. Section 3 introduces the heuristic switching model and conducts the analysis without the zero lower bound. In Section 4 we add the zero lower bound to the model and analyze liquidity traps. Simulations with policy interventions are presented in Section 5, and Section 6 concludes. In the appendix we present a micro-foundation of our heterogeneous expectations framework and proofs of the results.

2 Inflation targeting model

In order to facilitate comparison with the rational expectations benchmark we use a standard New Keynesian model in line with Galí (2002) and Woodford (2003). Micro foundations of this model under heterogeneous expectations are derived in appendix A. This derivation is closely related to Kurz et al. (2013), and makes use of the properties of our heuristic switching model, defined in Section 3. The New Keynesian Phillips curve and IS curve, describing inflation and output gap respectively, are given by

$$\pi_t = \beta E_t \pi_{t+1} + \kappa x_t + e_t, \tag{1}$$

$$x_t = E_t x_{t+1} + \frac{1}{\sigma}(E_t \pi_{t+1} - i_t) + u_t. \tag{2}$$

Here β is the discount factor, and

$$\kappa = \frac{(\sigma + \eta)(1 - \omega)(1 - \beta\omega)}{\omega}, \tag{3}$$

with σ and η the inverses of respectively the elasticity of intertemporal substitution and the elasticity of labor supply. $(1 - \omega)$ is the fraction of firms that can adjust their price in a given period, and i_t is the nominal interest rate, which can be freely chosen by the central bank. e_t and u_t are shocks to respectively inflation and output gap. Shocks to inflation can be interpreted as cost-push shocks. Shocks to output gap consist of changes in productivity.

We assume the central bank wants to reach its inflation target both now and in the future. It is furthermore assumed that the CB always wants output to be at its efficient level, which in our model corresponds to zero output gap. More specifically, the CB wants to minimize a loss function with the discounted sums of all squared deviations from these targets:

$$E_t \sum_{i=0}^{\infty} \beta^i \left[(\pi_{t+i} - \pi^T)^2 + \mu (x_{t+i})^2 \right]. \tag{4}$$

Here $\mu \geq 0$ is the relative weight that the central bank assigns to the minimization of the squared output gap compared to the squared deviation of inflation from target.

Rotemberg and Woodford (1999) and Woodford (2002) and others have shown that such a loss function can be derived from optimization of a second order approximation of the utility of a representative consumer. The optimal inflation target is then however implied to be 0. In this paper we analyze whether monetary policy aimed at minimizing the above loss function results in desirable dynamics under heterogeneous expectations, both with and without the restriction of $\pi^T = 0$.

There are two ways the CB could minimize the loss function. If the CB optimizes under discretion, it chooses π_t and x_t to minimize the loss function in every period with the current information. If the CB optimizes under commitment it commits to a policy rule now, and does not reconsider this rule in future periods. This way it can influence future private sector expectations and will therefore ultimately be better off. The main problem with this approach is that the

central bank will be tempted to re-optimize in every period. Commitment is only better for the CB because of the effect on private sector expectations. When those expectations have been formed the CB would be better off to renege on its commitment. However, the CB would than lose its credibility, so that we would be back in the discretion case.[3] In this paper we assume the CB optimizes under discretion. In that case, the first order conditions that are obtained from minimizing (4) result in the following optimal trade-off between inflation and the output gap:

$$\pi_t - \pi^T = -\frac{\mu}{\kappa} x_t. \tag{5}$$

The optimal policy rule that does not assume rational expectations and implements the above condition is derived by Evans and Honkapohja (2003). The same rule in slightly different settings is derived by Berardi and Duffy (2007), and Gomes (2006). Evans and Honkapohja (2003) study optimal monetary policy under learning, with non-rational, but homogeneous expectations. They find that this rule leads to convergence to the optimum under discretion even if expectations are not rational. Branch and Evans (2011) find that the rule also performs well in their model with heterogeneous expectations. The rule is given by

$$i_t = \psi_0 + \psi_1 E_t \pi_{t+1} + \psi_2 E_t x_{t+1} + \psi_3 u_t + \psi_4 e_t, \tag{6}$$

$$\psi_0 = -\frac{\sigma\kappa}{\mu + \kappa^2}\pi^T \qquad \psi_1 = 1 + \frac{\sigma\kappa\beta}{\mu + \kappa^2} \qquad \psi_2 = \psi_3 = \sigma \qquad \psi_4 = \frac{\sigma\kappa}{\mu + \kappa^2}$$

We assume that the central bank can perfectly observe private sector expectations, which is also done by the authors mentioned above, as well as by Branch and McGough (2010). Although the CB can respond to current period expectations, those expectations are assumed to be based on past information (as is standard in the learning literature). It is furthermore assumed that the central bank cannot respond to current period shocks. This way agents are not able to influence current period variables. Svensson (2003) strongly argues in favor of this.

In this paper we assume a more general interest rate rule and consider the optimal policy benchmark as a special case. We assume a forward looking Taylor rule that replaces contemporaneous values of inflation and output gap by expectations in the rule proposed by Taylor (1993).

[3] See e.g. Clarida et al. (1999) for further details.

Table 1: Calibrations and corresponding coefficients

Author(s)	$\sigma = \phi_2^{opt}$	κ	ϕ_1^{opt}	ϕ_1^{PF}	ϕ_1^{PD}
W	0.157	0.024	1.015	-5.607	20.56
CGG	1	0.3	1.874	-2.367	10.97
MN	1/0.164	0.3	6.326	-19.53	61.77

Since the long run real interest rate is 0 in our model, the Forward looking Taylor rue can be written as:

$$i_t = \pi^T + \phi_1(E_t\pi_{t+1} - \pi^T) + \phi_2 E_t x_{t+1}. \tag{7}$$

When we follow Rotemberg and Woodford (1999) and Woodford (2002) and assume that (in the absence of considerations for the zero lower bound on the nominal interest rate) welfare is maximized with a zero inflation target, optimal monetary policy is given by

$$\pi^T = 0, \qquad \phi_1^{opt} = 1 + \frac{\sigma\kappa\beta}{\mu + \kappa^2}, \qquad \phi_2^{opt} = \sigma \tag{8}$$

In the limit of β going to 1, the forward looking Taylor rule, (7) with optimal coefficient ϕ_1^{opt} and ϕ_2^{opt} can also be used to minimize (4) for nonzero inflation targets. When β is close to 1 (as is usually the case in calibrations), this rule gives a close approximation of optimal policy for $\pi^T \neq 0$.

In order to get an idea of the magnitude of the optimal policy parameters, the model needs to be calibrated. The first two columns of Table 1 give the calibrations of σ and κ of Woodford (1999), Clarida et al. (2000), and McCallum and Nelson (1999). Under all calibrations the discount factor is set to $\beta = 0.99$. Column 3 of Table 1 states the corresponding optimal values of ϕ_1. Here the weight on output gap, μ, is set to 0.25, as is done by McCallum and Nelson (2004), and Walsh (2003). Since under optimal policy $\phi_2 = \sigma$, the different optimal values of ϕ_2 can be read from Column 1 of Table 1. Column 4 and 5 are discussed in Section 3.

The Calibrations of Woodford (1999) and Clarida et al. (2000) result in reasonable optimal policy parameters that are close to empirical estimates of Taylor rules, both with contemporaneous values and with expectations of inflation and output gap.[4] The McCallum and Nelson (1999)

[4]See e.g Taylor (1999), Clarida et al. (2000) and Orphanides (2004).

calibration on the other hand gives rise to what seems to be extremely aggressive monetary policy.

Abstracting from shocks and plugging (7) into (2), gives the following model

$$x_t = (1 - \frac{\phi_2}{\sigma})E_t x_{t+1} - \frac{\phi_1 - 1}{\sigma}(E_t \pi_{t+1} - \pi^T), \tag{9}$$

$$\pi_t = \beta E_t \pi_{t+1} + \kappa x_t. \tag{10}$$

We assume here that the model parameters are positive, and that the policy parameters of the central bank are nonnegative.

Assumption 1. $\kappa, \sigma > 0, \;\; \phi_1, \phi_2, \pi^T \geq 0.$

3 Analysis with heuristic switching model

In this Section a heuristic switching model is used to analyze the dynamics of output gap and inflation when expectations are non-rational and heterogeneous. In a heuristic switching model as in Brock and Hommes (1997), beliefs are formed by a set of simple rules of thumb, or heuristics. The population consists of agents that can switch between those heuristics. As a heuristic preforms better in the recent past, the fraction of the population that follows that prediction rule increases. Agents are therefore learning over time by evolutionary selection based upon relative performance. The fractions of agents following the different heuristics evolve according to the following discrete choice model with multinomial logit probabilities (see Manski et al. (1981)):

$$n_{h,t} = \frac{e^{bU_{h,t-1}}}{\sum_{h=1}^{H} e^{bU_{h,t-1}}}. \tag{11}$$

Here $n_{h,t}$ is the fraction of agents that follows heuristic h in period t, and $U_{h,t}$ is the fitness measure of heuristic h in period t, i.e., a measure of how well the heuristic performed in the past. Finally, b is the intensity of choice. The higher the intensity of choice, the more sensitive agents become with respect to relative performance of the heuristics.

We assume private sector beliefs are formed by two simple, but plausible heuristics: fundamentalists and naive. Followers of the naive heuristic make use of the high persistence in inflation and

output gap dynamics, and believe future inflation or output gap to be equal to their last observed values. Note that the naive forecast is optimal when inflation and output follow a random walk, and close to optimal when the system contains a near unit root.

Followers of the fundamentalist heuristic on the other hand believe inflation or output gap to be equal to the fundamental values that would arise under rational expectations. Fundamentalists thus act as if all agents are rational. They do not take into account that there are other agents in the economy, as they lack the cognitive ability to know exactly the beliefs of other agents or the number of agents with different expectations. However, as long as other agents make the same predictions as the fundamentalists (not necessarily by using the same heuristic), fundamentalists will have perfect foresight in the absence of shocks.

In order to be able to assess the credibility of the central bank, we will only consider specifications of our model in which rational expectations (and thus the expectations of fundamentalists) coincide with the targets of the central bank. Fundamentalists expectations are then equal to $x_t = 0$ and $\pi_t = \pi^T$, and their expectations could alternatively be interpreted as having been formed by trusting the central bank. Whichever value of inflation the central bank targets, fundamentalist believe that the central bank will be able to achieve it, so that they expect future inflation to be equal to this target.

Note that we talk about credibility of the central bank's targets values of inflation and output gap, and not about the credibility of its future policy actions (as credibility is often referred to in the literature). This is in line with the fact that inflation targeting can be seen as a commitment to goals rather than a commitment to the CB's future actions and details of its operations. Credibility over targets furthermore implicitly captures both the CB's willingness to take actions to achieve its targets *and* its ability to do so, where the latter is not straightforward in an economy with boundedly rational agents.

Since in our model it is possible in any given period that one heuristic performs better in forecasting output gap while the other performs better in forecasting inflation, we do not impose any ex ante constraints on the relation between output gap and inflation that agents expectations must satisfy. Instead we allow the fraction of fundamentalists, denoted n_t^z, to differ between

inflation ($z = \pi$) and output gap ($z = x$).[5] Agents will then learn to use the best heuristic for each variable. This can be the same heuristic in times where the time series of inflation and output gap have similar features. However, in periods of hyper inflation with an output gap close to zero, agents will learn to be fundamentalistic about the output gap, but to use past inflation as a best predictor of future inflation. Furthermore, this set up of the model also allows for periods where the CB is perfectly credible in its inflation fighting policy, but where agents do not believe it will be able to keep the output gap at zero at the same time.

Finally, let the fitness measure for both variables be a weighted sum of the negative of the last observed squared prediction error, and the previous value of the fitness measure.

$$U^z_{t-1} = -(1-\rho)(z_{t-1} - E_{t-2}z_{t-1})^2 + \rho U^z_{t-2}, \qquad z = \pi, x, \tag{12}$$

where $0 \leq \rho \leq 1$, is the memory parameter. For analytical tractability we set $\rho = 0$ for now, and reintroduce the parameter in the simulations in Section 5.

To simplify calculations and presentation we introduce a new variable which is defined as the difference between the fraction of fundamentalistic agents (n^z_t) and the fraction of naive agents (1-n^z_t).

$$m^z_t = n^z_t - (1 - n^z_t) = 2n^z_t - 1, \qquad z = \pi, x. \tag{13}$$

When all agents are fundamentalistic the difference in fractions equals 1, and when all agents are naive, the difference in fractions equals -1. Henceforth we will refer to these differences in fractions simply as fractions. We can interpret these fractions as *endogenous credibility*. When $m^x_t = m^\pi_t = 1$ the central bank has full credibility, and when $m^x_t = m^\pi_t = -1$ the CB has lost all its credibility. This credibility measure will turn out to be very important in determining the effectiveness of monetary policy.

Using (13), average expectations about inflation and output gap can be written as

$$E_t\pi_{t+1} = \frac{(1 + m^\pi_t)}{2}\pi^T + \frac{(1 - m^\pi_t)}{2}\pi_{t-1}, \tag{14}$$

[5] All results presented in this section continue to hold when we impose that n^x_t and n^π_t should evolve together. Results presented in Section 4 and 5 will also remain valid qualitative.

$$E_t x_{t+1} = \frac{(1 - m_t^x)}{2} x_{t-1}. \tag{15}$$

Using these expectations, (10) and (9) can be written as

$$x_t = (1 - \frac{\phi_2}{\sigma}) \frac{(1 - m_t^x)}{2} x_{t-1} - \frac{\phi_1 - 1}{\sigma} \frac{(1 - m_t^\pi)}{2} (\pi_{t-1} - \pi^T), \tag{16}$$

$$\pi_t = \beta \frac{(1 + m_t^\pi)}{2} \pi^T + \beta \frac{(1 - m_t^\pi)}{2} \pi_{t-1} + \kappa x_t. \tag{17}$$

To complete the model we specify m_{t+1}^x and m_{t+1}^π by combining (11), (12) and (13). This gives

$$m_{t+1}^x = \text{Tanh}\left(\frac{b}{2} (x_{t-2}^2 - 2 x_t x_{t-2}) \right), \tag{18}$$

$$m_{t+1}^\pi = \text{Tanh}\left(\frac{b}{2} (\pi_{t-2}^2 - (\pi^T)^2 - 2(\pi_{t-2} - \pi^T)\pi_t) \right). \tag{19}$$

The above system is six dimensional. First of all, next periods inflation and output gap (π_{t+1} and x_{t+1}) are determined by the current values of these variables (π_t and x_t), and by next periods fractions (m_{t+1}^π and m_{t+1}^x). These four variables are however not enough to determine the future dynamics of the system since m_{t+2}^π and m_{t+2}^x, which determine π_{t+2} and x_{t+2}, depend on π_{t-1} and x_{t-1}, and are therefore not determined by the above mentioned variables. It follows that the system must be six dimensional and that the state vector can be written as

$$\begin{pmatrix} x_t & \pi_t & x_{t-1} & \pi_{t-1} & m_{t+1}^x & m_{t+1}^\pi \end{pmatrix}, \tag{20}$$

or as

$$\begin{pmatrix} x_t & \pi_t & m_{t+2}^x & m_{t+2}^\pi & m_{t+1}^x & m_{t+1}^\pi \end{pmatrix}. \tag{21}$$

3.1 Steady states and stability

The central bank aims to keep inflation at its target level and output gap at zero. It would therefore be desirable for our dynamical system to have a steady state with $\pi = \pi^T$ and $x = 0$. Proposition 1 states that such a steady state indeed exists either when the inflation target is

zero, or in the limit of the discount factor going to unity. The proof of Proposition 1 is given in Appendix B.1.

Proposition 1. *When either* $\pi^T = 0$ *or* $\beta \to 1$*, a steady state with* $x^* = 0, \pi^* = \pi^T$*,* $m^{x*} = 0, m^{\pi*} = 0$ *exists.*

Since the steady state with $x^* = 0, \pi^* = \pi^T$ coincides with the rational expectations equilibrium values of our model, we call this steady state the *fundamental steady state*. Even though in this steady state convergence to rational expectation *values* has taken place, it is not the case that all agents use the fundamentalist *heuristic*. This is so because the naive heuristic also gives perfect steady state predictions, so that both fundamentalists and naive agents have perfect foresight at the fundamental steady state. The difference in fractions therefore equals zero for both variables ($m^{x*} = m^{\pi*} = 0$).

As shown in Appendix B.1, when β is close to 1 (e.g. 0.99) the fundamental steady state lies close to the values given in Proposition 1. This will also been shown to be true in simulations in Section 5. In the remainder of this section we will both consider the case of $\pi^T = 0$, and the case of $\pi^T > 0$ with $\beta \to 1$. The latter case gives an approximation of how results change when a general inflation target is chosen.

The central bank would like to achieve convergence to the fundamental steady state from as wide a range of initial conditions as possible. This requires first of all that the fundamental steady state is *locally stable*. The central bank can try to achieve stability of the fundamental steady state by choosing the right values of the parameters in its monetary policy rule, ϕ_1 and ϕ_2. The inflation target π^T turns out not to matter for stability of the fundamental steady state.

In order for the steady state to be locally stable, it is required that all six eigenvalues are inside the unit circle at the steady state. In Appendix B.2 it is shown that in the fundamental steady state four eigenvalues are equal to zero and that the other two eigenvalues are given by

$$\lambda_1 = \frac{1}{4}\left((1+\beta-\frac{\phi_2}{\sigma}-\kappa\frac{\phi_1-1}{\sigma}) + \sqrt{\left(1+\beta-\frac{\phi_2}{\sigma}-\kappa\frac{\phi_1-1}{\sigma}\right)^2 - 4\beta(1-\frac{\phi_2}{\sigma})}\right), \tag{22}$$

and

$$\lambda_2 = \frac{1}{4}\left((1+\beta-\frac{\phi_2}{\sigma}-\kappa\frac{\phi_1-1}{\sigma}) - \sqrt{\left(1+\beta-\frac{\phi_2}{\sigma}-\kappa\frac{\phi_1-1}{\sigma}\right)^2 - 4\beta(1-\frac{\phi_2}{\sigma})}\right). \tag{23}$$

When $\lambda_1 = 1$ or $\lambda_2 = -1$ a bifurcation occurs, and the fundamental steady state loses its stability. Proposition 2 and 3 describe when this occurs. These results are illustrated in Figure 1 and will be discussed below. The proofs of the propositions are given in Appendix B.3 and B.4 respectively.

Proposition 2. *(See Figure 1) When $\pi^T = 0$ and the CB chooses*

$$\phi_1 < \phi_1^{PF} = 1 - (2-\beta)\frac{\sigma+\phi_2}{2\kappa}, \tag{24}$$

the fundamental steady state is unstable due to a subcritical pitchfork bifurcation (with two unstable, non-fundamental steady state above the bifurcation value), with $\lambda_1 = +1$. The bifurcation value for a nonzero inflation target can be obtained by letting $\beta \to 1$ in (24)*.*

Proposition 3. *(See Figure 1) When $\pi^T = 0$ and the CB chooses*

$$\phi_1 > \phi_1^{PD} = 1 + (2+\beta)\frac{3\sigma-\phi_2}{2\kappa}, \tag{25}$$

the fundamental steady state is unstable due to a period doubling bifurcation, with $\lambda_2 = -1$. This bifurcation is subcritical (with a 2-cycle below the bifurcation value) if $\phi_2 < 3\sigma$ and supercritical (with a 2-cycle above the bifurcation value) if $\phi_2 > 3\sigma$. The bifurcation value for a nonzero inflation target can be obtained by letting $\beta \to 1$ in (25)*.*

It follows from Proposition 2 and 3 that the fundamental steady state can either be unstable because the central bank responds too weakly or because the CB responds too strongly to inflation and output gap expectations.

The intuition of instability of the fundamental steady state due to monetary policy that reacts too weakly is the following. If period t expectations of period $t+1$ inflation and/or output gap are high, and the central bank does not respond with a large enough increase in the interest rate, these high expectations will lead to period t realizations of inflation and output gap that are even higher. This will lead expectations about period $t+2$, formed in period $t+1$, to be

higher than expectations about period $t+1$, leading to even higher period $t+1$ realizations. This will lead to a loss of credibility for the central bank (more agents become naive), leading to more instability. What follows is a continued rise of both inflation and output gap, together with declining credibility and rising expectations: an inflationary spiral. Analogously, for low initial conditions a deflationary spiral will occur under weak monetary policy

If the central bank responds too strongly to expectations, high inflation and/or output gap expectations about period $t+1$ are countered in period t by a very high interest rate. This results in very low inflation and output gap realizations in period t, leading to very low expectations about period $t+2$. The CB then sets the interest rate very low in period $t+1$, leading to even higher inflation and output gap realizations in period $t+1$ than agents had expected. This causes a loss in credibility, and high naive expectations about period $t+3$. The following increase in the interest rate leads to even lower realizations in period $t+2$ than agents expected, again leading to a loss in credibility and more extreme expectations. These cyclical dynamics continue, leading inflation and output gap to jump up and down between ever higher and lower values: explosive overshooting.

The results of Proposition 1, Proposition 2 and Proposition 3 can be combined in a bifurcation diagram of ϕ_1. This is done in Figure 1 for the case of $\beta \to 1$, with ϕ_1 on the horizontal axis and π_t on the vertical axis. The fundamental steady state is located at $\pi_t = \pi^T$, and the black line between ϕ_1^{PF} and ϕ_1^{PD} indicates the range of ϕ_1 values for which this steady state is (locally) stable. To the left of ϕ_1^{PF} and to the right of ϕ_1^{PD} the fundamental steady state is unstable, which is indicated by blue dashed lines. In this picture it is assumed that $\phi_2 = \sigma$, so that the period doubling bifurcation is subcritical. This implies the existence of an unstable 2-cycle to the *left* of ϕ_1^{PD}, which is depicted by the red dashed curves. The blue dashed curves between ϕ_1^{PF} and 1, represent the non-fundamental unstable steady states from Proposition 1 that are created in the subcritical pitchfork bifurcation. As discussed above, explosive cyclical dynamics, due to overshooting, occur to the right of ϕ_1^{PD}. To the left of ϕ_1^{PF} inflation either monotonically increase or decrease, depending on initial conditions. In Figure 1 it is assumed that initial output gap is zero, so that the inflation target is the boundary between inflationary and deflationary spirals.

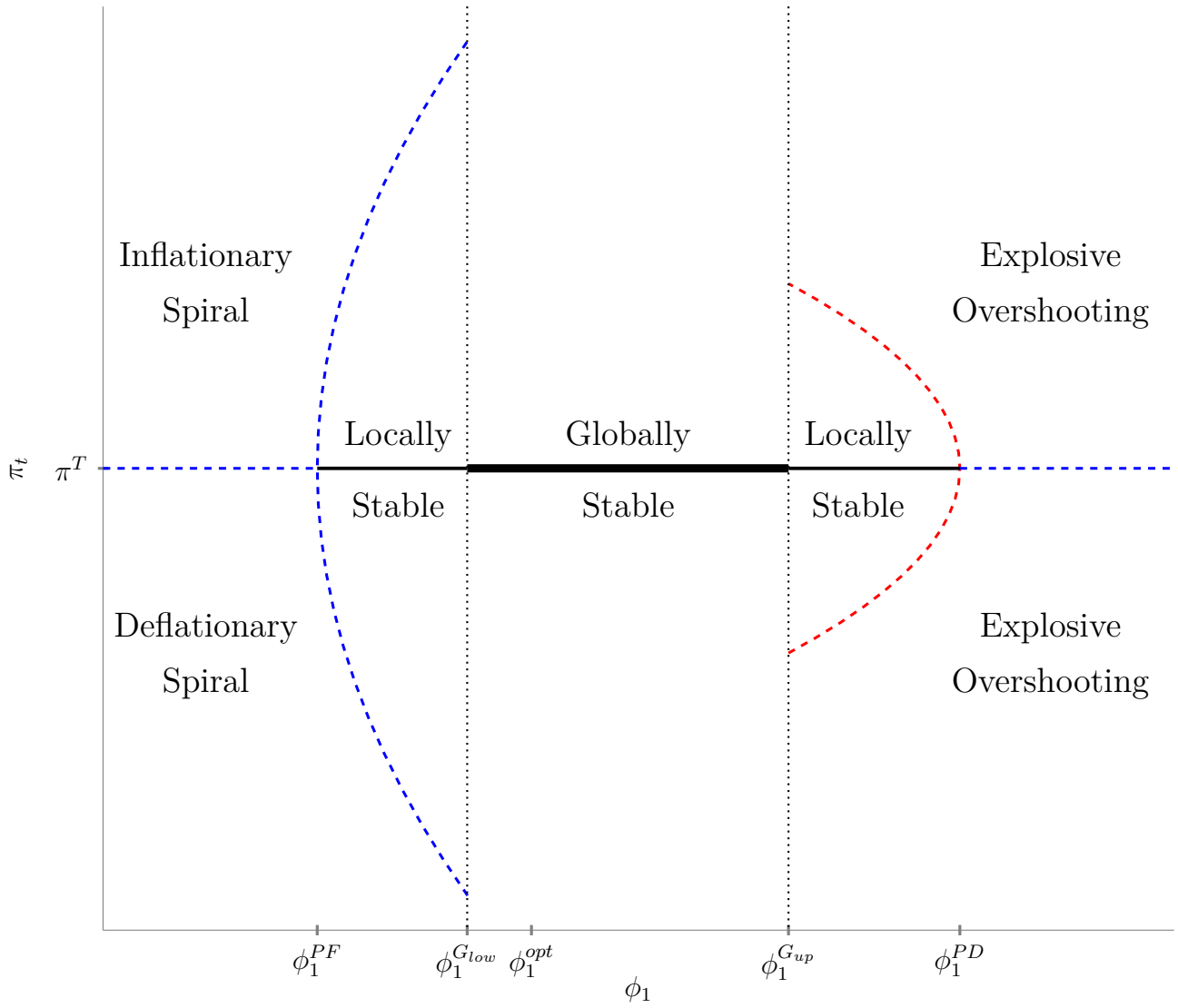

Figure 1: Bifurcation diagram of ϕ_1 in case of $\phi_2 = \sigma$ and $\beta \to 1$. The locally (globally) stable area of the fundamental steady state is indicated with a solid (thick) black line. Unstable steady states are indicated with blue dashed curves, while the red dashed curves represent an unstable 2-cycle. When the fundamental steady state is not globally stable, explosive overshooting can occur if monetary policy is too aggressive. When monetary policy is too weak, either a inflationary or a deflationary spiral will occur. In this picture of the dynamics it is assumed that initial output gap is at its steady state level.

The result of Proposition 2 and Proposition 3 are similar to the conditions required for local determinacy under rational expectations found by Bullard and Mitra (2002). These authors show that, when the central bank responds to inflation and output gap expectations, determinacy of the rational expectations equilibrium requires both that the well known Taylor principle is satisfied, and that the central bank does not to respond too strongly to expectations.

More specifically the authors find that $\phi_1 > 1 - (1-\beta)\frac{\phi_2}{\kappa}$ must hold. However, the condition for local stability that follows from Proposition 2 requires ϕ_1 to be larger than ϕ_1^{PF}, which is (under Assumption 1) strictly smaller than the value found by Bullard and Mitra (2002). We can therefore have local stability even if the Taylor principle is not satisfied. The second condition for determinacy given by Bullard and Mitra (2002) reduces to $\phi_1 < 1 + (1+\beta)\frac{2\sigma-\phi_2}{\kappa}$. This condition is again strictly stronger than our condition for local stability, which requires $\phi_1 < \phi_1^{PD}$.[6]

We can conclude that with heterogeneous expectations the range of policy parameters that are allowed in order to have a locally stable fundamental steady state is strictly larger (in both directions) than the range of parameters allowed under rational expectations in order to have a locally determinate equilibrium. However, the conditions for a locally determinate rational expectations equilibrium coincide with the conditions for stability under all naive expectations. Since all naive expectations is the most unstable case in our model, these conditions imply *global* stability of the fundamental steady state in our model. This is stated in Proposition 4, the proof of which is given in Appendix B.5.

Proposition 4. *If $\pi^T = 0$, the fundamental steady state is globally stable when the central bank chooses*

$$1 - (1-\beta)\frac{\phi_2}{\kappa} = \phi_1^{G_{low}} < \phi_1 < \phi_1^{G_{up}} = 1 + (1+\beta)\frac{2\sigma - \phi_2}{\kappa}, \tag{26}$$

The global stability conditions for a nonzero inflation target can be obtained by letting $\beta \to 1$ in (26).

[6] We also analyzed local stability under a more traditional Taylor rule where the central bank responds to contemporaneous values of inflation and output. Here we find that local stability of the fundamental steady state requires that $\phi_1 > \frac{1}{2}(1 - (2-\beta)\frac{\sigma+2\phi_2}{2\kappa})$, which is a strictly weaker condition than the one found in Proposition 2. Furthermore, just as under rational expectations, there is no upper bound on the monetary policy parameters under a contemporaneous Taylor rule.

The global stability region of the fundamental steady state is indicated by the thick black line in Figure 1. For this region of policy parameters no unstable steady states or 2-cycles exist.

3.2 Policy implications

The difference between the local and the global stability results of the previous section highlight the importance of the credibility of the central bank in stabilizing the economy. When the central bank is able to retain a substantial amount of credibility, even after a sequence of bad shocks, its conditions on monetary policy will not be very restrictive and lie close to those given in Proposition 2 and 3. In our model, this situation would e.g. occur if the intensity of choice is not too high. If, on the other hand, the central bank is likely to lose *all* its credibility after a sequence of bad shock, the restrictions on monetary lie close to those given in Proposition 4. This is in line with the results of the agent based model of Salle et al. (2013), who find that under low (exogenous) credibility of the central banks inflation target, conditions on policy parameters are much more restrictive than under high credibility. As in our model, the Taylor principle is furthermore not a necessary condition in the latter case.

An important question now is whether the fundamental steady state is locally, or perhaps even globally stable under the theoretically *optimal* monetary policy that is implemented by choosing parameter values as in (8). Proposition 5 states that this is indeed the case. Its proof is given in B.6

Proposition 5. *When the central bank implements optimal policy by choosing the values for* π^T, ϕ_1 *and* ϕ_2 *from* (8)*, the fundamental steady state is globally stable.*

In Figure 1 it is indicated that ϕ_1^{opt} lies in the globally stable area where the non-fundamental steady states and 2-cycle do not exist. It follows that the theoretically optimal policy parameters are indeed a desirable choice for the central bank in our model with heterogeneous expectations.

It is however also of interest to know whether small deviations from optimal policy can lead to instability. That is, whether or not the bifurcations values from Proposition 2 and 3 lie close to the optimal parameter setting, or the bifurcations occur for parameter values that a central bank would never choose in practice. To investigate this we look at the calibrations discussed in Section

2. Column 4 and 5 of Table 1 state the values of the pitchfork bifurcation (ϕ_1^{PF}) and the period doubling bifurcation (ϕ_1^{PD}), given that ϕ_2 is chosen optimally.

From Column 4 it follows that the pitchfork bifurcation occurs at negative values for all calibrations. This means that, in contrast to the Taylor principle, under these calibrations the fundamental steady state is locally stable for monetary policy that reacts weakly to inflation ($0 < \phi_1 < 1$) as long as ϕ_2 is chosen optimally. This result furthermore turns out to hold for any nonnegative choice of ϕ_2.

The values of the period doubling bifurcation (ϕ_1^{PD}) given in Column 5 of Table 1 are all unrealistically high. This means that when ϕ_2 is chosen optimally, reacting too strongly to inflation will not be a problem for any reasonable value of ϕ_1. The dependence of this result on the optimality of the output gap coefficient however drastically differs over calibrations. Under the Clarida et al. (2000) calibration a very strong output gap coefficient of $\phi_2 = 2$ results in instability for $\phi_1 > 6$, implying that responding too aggressively will not be a problem. However, under the Woodford (1999) calibration, an output gap parameter of $\phi_2 > 0.49$ implies that the period doubling bifurcation occurs at a negative value of ϕ_1, so that the fundamental steady state will be unstable for *any* positive inflation parameter.

4 Zero lower bound on the interest rate

In the previous section no restrictions were placed on the values that can be taken by the nominal interest rate. In practice, the nominal interest rate will however never be set negative. We will show that if the zero lower bound (ZLB) is accounted for, the global stability result of Proposition 4 no longer holds, but that with the ZLB deflationary spirals with ever decreasing inflation and output gap can always arise, even under optimal policy. We show this in a sequence of propositions for the limiting case of infinite intensity of choice, i.e., when all agents immediately switch to the best predictor, in Section 4.1. In Section 4.2 we argue that for finite intensity of choice qualitatively similar dynamics occur.

First we show that the introduction of the ZLB can lead to the existence of an additional steady state (Proposition 6). The appearance of this additional steady state (or equilibrium) was first

highlighted by Benhabib et al. (2001*a*,*b*) under rational expectations. The presence of this steady state causes divergence to minus infinity for low inflation and output gap in our model, just as in Evans et al. (2008) and Benhabib et al. (2014). Whether such a deflationary spiral occurs however not only depends on initial inflation and output gap, but also on the *credibility* of the central bank (i.e., the fractions of fundamentalists). We argue that a liquidity trap can never arise as long as the CB retains full credibility (Proposition 7), and that a self-fulfilling deflationary spiral only arises when naive agents perform better than fundamentalists about both variables (Proposition 8). However, even in a liquidity trap where the CB has lost *all* its credibility, recovery is still possible if inflation and output gap are not too low. The deflationary spiral and recovery regions are illustrated in Figure 2. The corresponding sufficient conditions for recovery or a deflationary spiral to occur are found in Proposition 9. This proposition also shows that initial conditions for which a deflationary spiral occurs always exist.

Our model is capable of describing expectation driven liquidity traps. We can interpret low initial inflation and output gap as having been caused by a negative shock to the fundamentals of the economy. Under rational expectations the economy would immediately recover from such a shock if there are no new negative shocks in the next period. This is not the case in our model of heterogeneous expectations. Here it is likely that the low realizations of inflation and output gap caused by the shock, lead to a loss of credibility of the central bank, i.e., a higher fraction of naive agents. These naive agents expect the low realizations of inflation and output gap, and therefore the liquidity trap, to continue. These low expectations then lead to low realizations of inflation and output gap in the next period, so that the liquidity trap indeed continues, and expectations become self-fulfilling. If the shocks to inflation and output gap were not too large, or if the central bank retained enough credibility, both variables will start to rise again, and recovery to the positive interest rate region occurs. However, if expectations are too low, inflation and output gap start to decline, resulting in more loss of credibility. The economy then ends up in a self-fulfilling deflationary spiral with no credibility for the central bank.

4.1 Analysis for infinite intensity of choice

When we introduce the zero lower bound, the interest rate becomes piecewise linear. In normal times the interest rate is still given by Equation (7), but when this equation implies that the nominal interest rate is negative, it is instead set equal to zero. This happens when

$$E_t\pi_{t+1} + \frac{\phi_2}{\phi_1}E_t x_{t+1} < (1 - \frac{1}{\phi_1})\pi^T. \tag{27}$$

The model that results is still described by Equation (16) through (19) when Equation (27) is not satisfied. We will refer to combinations of expectations where this is the case as the "positive interest rate region". Combinations of expectations where (27) holds will be referred to as the "zero lower bound region", or simply the "ZLB region". In the ZLB region the model is described by

$$x_t = \frac{(1 - m_t^x)}{2}x_{t-1} + \frac{(1 + m_t^\pi)}{2\sigma}\pi^T + \frac{(1 - m_t^\pi)}{2\sigma}\pi_{t-1}, \tag{28}$$

$$\pi_t = \beta\frac{(1 + m_t^\pi)}{2}\pi^T + \beta\frac{(1 - m_t^\pi)}{2}\pi_{t-1} + \kappa x_t, \tag{29}$$

with fractions given by (18) and (19). The steady states of this nonlinear system depend on the fractions of agents following the different heuristics and therefore are, in general, quite difficult to analyze. For this reason we first consider the limiting case where the intensity of choice equals infinity, and all agents immediately switch to the best performing heuristic. The (6-dimensional) system then becomes piecewise linear.

Furthermore, to make an analysis of the zero lower bound interesting in this setup, we assume a positive inflation target ($\pi^T > 0$). For reasons discussed above, this requires us to consider, in the remainder of this section, the limiting case of $\beta \to 1$. However, as was the case in the previous section, considering a discount factor of e.g. 0.99, would only marginally change the results.

Proposition 6 states that, when the intensity of choice equals infinity, at most one (unstable) steady state exists in the ZLB region. In this steady state fundamentalists make persistent prediction errors about inflation, implying that *all* agents have switched to naive expectations about this variable. Naive agents do not make prediction errors, so in this steady state expectations are

perfectly self-fulfilling.

Proof of Proposition 6 is given in Appendix C.1.

Proposition 6. *For $b = +\infty$ there exists exactly one steady state in the ZLB region when the Taylor principle is adhered to ($\phi_1 > 1$). This steady state is defined by $x = 0$, $\pi = 0$, $m^x = 0$ and $m^\pi = -1$, and always is an unstable saddle point. When the Taylor principle is not adhered to ($\phi_1 < 1$) no steady states exist in the ZLB region.*

Initial conditions in the ZLB region typically will not lead to convergence to the steady state of Proposition 6 since it is unstable. Two generic possibilities that can occur for initial conditions in the ZLB region are the following. First of all, it is possible that inflation and output gap start increasing, and at some point cause the system to cross the zero lower bound and enter the positive interest rate region. From then on the dynamics will be as in Section 3. That is, for monetary policy that satisfies the conditions of Proposition 4 (and under some additional conditions also policy that satisfies Proposition 2 and 3) convergence to the fundamental steady state occurs. The other possibility for dynamics in the ZLB region is that inflation and output gap decline towards minus infinity. Such a deflationary spiral can be interpreted as an inescapable liquidity trap. We will refer to the first case as "recovery", and to the second case as "divergence", or as a "deflationary spiral".

When the intensity of choice equals infinity all fractions are either -1, 1, or 0. Fractions of 0 will typically only occur in a steady state and are not relevant for out-of-steady-state dynamics. This possibility will therefore not be considered below. It is convenient to use (21) as state vector because four of the six variables will take only two different values. These four state variables can be represented by a table of 16 different combinations. This is done in Table 2, illustrating which initial conditions lead to recovery and which to divergence.

In Proposition 7 it is stated that for the four cases in the first column of Table 2 initial fractions are such that the system already is in the positive interest rate region from the very first period onwards. For the cases in the first row of Table 2 the system trivially is in the positive interest rate region after one period. Therefore it is indicated in Table 2 that for these cases recovery "always" occurs. The intuition behind this result is that the economy can never be in a liquidity

Table 2: Conditions for recovery for different initial conditions when $b = +\infty$

	$m^{\pi}_{t+1} = 1$		$m^{\pi}_{t+1} = -1$	
	$m^{x}_{t+1} = 1$	$m^{x}_{t+1} = -1$	$m^{x}_{t+1} = 1$	$m^{x}_{t+1} = -1$
$m^{\pi}_{t+2} = 1$, $m^{x}_{t+2} = 1$	Always	Always	Always	Always
$m^{\pi}_{t+2} = 1$, $m^{x}_{t+2} = -1$	Always	-	-	-
$m^{\pi}_{t+2} = -1$, $m^{x}_{t+2} = 1$	Always	-	$\pi_t > 0$	-
$m^{\pi}_{t+2} = -1$, $m^{x}_{t+2} = -1$	Always	-	$\pi_t > 0$	Deflationary spiral case (conditions in Proposition 9)

trap when the central bank has full credibility.

Proposition 7. *If at any point in time all agents are fundamentalistic about both inflation and output gap ($m^{\pi}_t = 1$ and $m^{x}_t = 1$; full credibility), recovery has occurred.*

Proof. When expectations about both variables are fundamentalistic we have $E_t\pi_{t+1} = \pi^T$ and $E_t x_{t+1} = 0$, and (27) can never hold for a nonnegative inflation target. Therefore, the system is in the positive interest rate region by definition. □

For the remaining nine cases of Table 2 recovery or divergence occurs conditional on the initial conditions of the other two state variables: π_t and x_t. For most cases it is not straightforward to define exactly for which values of π_t and x_t recovery and divergence occur. However, if a deflationary spiral occurs this must be because all agents have negative naive expectations about both variables after a few periods. That is, as long as the CB retains some credibility the economy has not (yet) entered a deflationary spiral. This is stated more formally in Proposition 8, the proof of which is given in Appendix C.2.

Proposition 8. *A necessary condition for a deflationary spiral to occur, is that at some point in time, $s \geq t$, all agents are naive with respect to both inflation and output gap for the next two periods ($m^{\pi}_{s+1} = -1$, $m^{x}_{s+1} = -1$, $m^{\pi}_{s+2} = -1$ and $m^{x}_{s+2} = -1$)*

From Proposition 8 it follows that a necessary condition for a deflationary spiral is that the system at some point has moved to the bottom right entry of Table 2, with all naive expectations. This entry is therefore the most interesting case when a deflationary spiral is concerned, and hence is labeled the "deflationary spiral case".

From any other entry in Table 2 either recovery occurs, or the system moves to the deflationary spiral case, after which the occurrence of recovery or divergence depends on the conditions of that case. We therefore do not present individual conditions for all these cases, but instead focus on the deflationary spiral case. Some intuition in when recovery or divergence might occur in other cases is presented in two lemmas in Appendix C.3. From these lemmas the necessary and sufficient conditions for recovery for the bottom two entries of the third column of Table 2 immediately follow.

We now turn to the deflationary spiral case. If here initial inflation and output gap are too low, expectations will remain naive and output gap and inflation will keep decreasing without bound. If, however, initial inflation and output gap are high enough, recovery occurs, either because of positive naive expectations, or because at some point expectations become fundamentalistic.

In Proposition 9 sufficient conditions for both recovery and divergence for the deflationary spiral case are presented. This proposition thereby also proofs that it is always possible to find initial conditions that lead to a deflationary spiral in our model. The proof of Proposition 9 is presented in Appendix C.4, and the corresponding deflationary spiral and recovery regions will be presented in Figure 2.

Proposition 9. *(See Figure 2) If all agents' expectations about both inflation and output gap are naive for two consecutive periods ($m^{\pi}_{t+1} = m^{\pi}_{t+2} = m^{x}_{t+1} = m^{x}_{t+2} = -1$) a sufficient condition for divergence to minus infinity is that*

$$x_t < -\max\left(\frac{1+\sqrt{1+4\frac{\sigma}{\kappa}}}{2\sigma}, \frac{4\sigma+2\kappa}{2\sigma\kappa+\sigma^2}\right)\pi_t \tag{30}$$

and either $\pi_t > \pi^T$ or

$$\pi_t < \frac{\sigma^2\pi^T - (4\kappa\sigma^2 + 2\kappa^2\sigma)x_t}{\sigma^2 + 6\kappa\sigma + 2\kappa^2} \tag{31}$$

This implies that for infinite intensity of choice a deflationary spiral can always occur if initial conditions are low enough.

A sufficient condition for initial output gap and inflation to lead to recovery is

$$x_t > -\max\left(\frac{1+\sqrt{1+4\frac{\sigma}{\kappa}}}{2\sigma}, \frac{4\sigma+2\kappa}{2\sigma\kappa+\sigma^2}\right)\min(\pi_t, \pi^T) \tag{32}$$

In Proposition 9 it is stated that the following three conditions are sufficient for divergence from all naive initial conditions. First of all, initial inflation and output gap must lie below the stable eigenvector through the steady state of Proposition 6 of the system with all naive expectations. If this condition (represented by the first part of the max function of (30)) is satisfied, divergence would always occur if agents were not allowed to become fundamentalistic. Secondly, initial conditions must be such that agents do not become fundamentalistic about output gap. This is the second part of the max function of (30). Finally, if inflation is below its target, agents must not become fundamentalistic about inflation (Condition (31)).

Analogously, as long as inflation is below its target, the following condition is sufficient for recovery: either initial inflation and output gap lie above the stable eigenvector of the all naive system, or agents become fundamentalistic about output gap. This condition explains the max function of (32). When initial inflation lies above its target fundamentalistic expectations could reduce inflation expectations to π^T. In that case a sufficient condition for recovery is that recovery would occur if inflation expectations were fundamentalistic. This reasoning explains the min function of (32).

Figure 2 plots the conditions from Proposition 9 in the (π, x)-plane for the Woodford (1999) calibration. The thick red line indicates the naive expectations zero lower bound for optimal policy with a weight on output gap of $\mu = 0.25$ and an annualized inflation target of 2%. This line separates the positive interest rate region from the ZLB region. Under the above calibration (31) (not plotted) is always satisfied when $\pi_t < \pi^T$ and (30) is satisfied. It therefore follows from Proposition 9 that a deflationary spiral (divergence) occurs for all initial conditions to the left of the steepest of the two sloped lines from condition (30), which is the dashed line in Figure 2. This line indicates the condition for agents to become fundamentalistic about output gap. The thin sloped line depicts the stable eigenvector of the all naive system through the steady state of

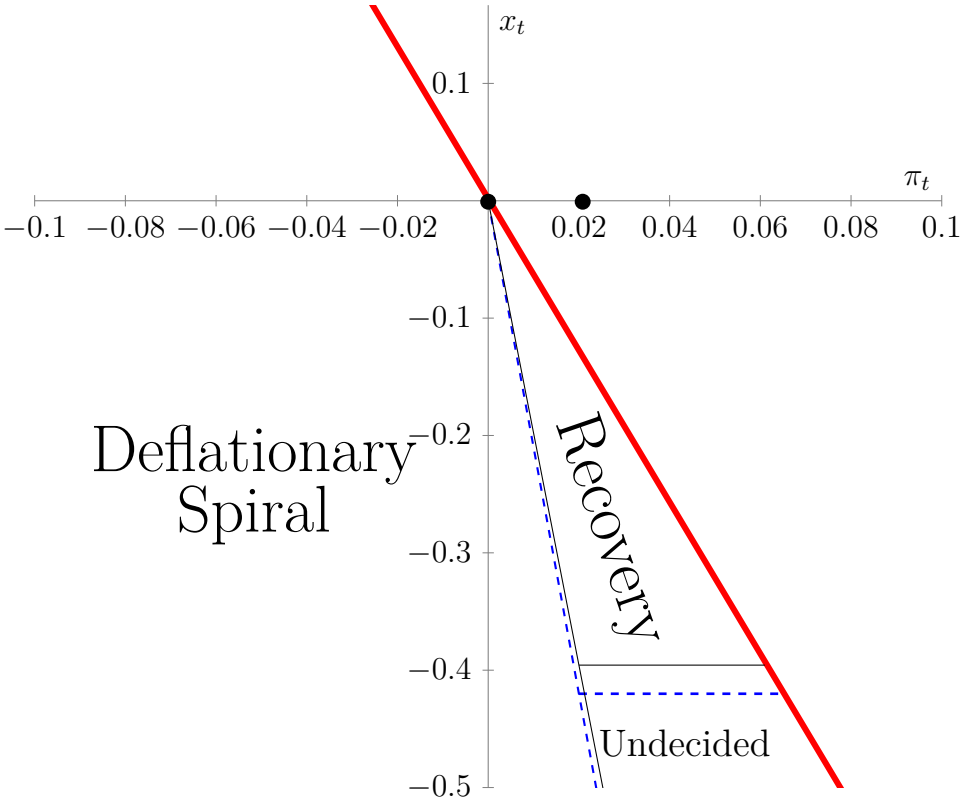

Figure 2: Conditions for recovery and divergence, presented in Proposition 9, for the Woodford (1999) calibration in the annualized $(\pi,\ x)$-plane. Here $b = +\infty$ and $m^x_{t+2} = m^\pi_{t+1} = m^\pi_{t+2} = m^x_{t+1} = -1$. The thick red line indicates the ZLB for naive expectations. The black dot at 2% inflation indicates the target steady state, while the black dot just below the ZLB depicts the unstable saddle steady state. The stable eigenvector through this saddle is depicted by the thin sloped line. For initial conditions to the left of the sloped dashed line agents will never become fundamentalistic about output gap, and a deflationary spiral occurs. For initial conditions to the right of this dashed line and above the horizontal dashed line agents will become fundamentalistic about output gap, and recovery occurs. The area below the horizontal dashed line is undecided since fundamentalistic expectations about inflation might prevent recovery here.

Proposition 9, which in turn is indicated by the black dot just below the ZLB. Recovery to the positive interest rate region occurs for initial conditions to the right of the sloped dashed line and above the horizontal dashed line, which arises due to the min function of (32). If monetary policy is neither too strong nor too weak, i.e., if it satisfies the conditions from Section 3.1, the economy will subsequently converge to the fundamental steady state, depicted by the other black dot.

The only area that is left indecisive by Proposition 9 is the area below the horizontal dashed line and to the right of the sloped dashed line. This area is however not very relevant since it consists of negative annualized output gaps of over 40%.

From the difference in scale on the axes of Figure 2 and the position of the recovery region, we can conclude that under this calibration inflation expectations are much more important than output gap expectations in determining whether recovery or divergence occurs in a liquidity trap. We find similar results under the Clarida et al. (2000) calibration.

4.2 Finite intensity of choice

Now we turn to the more general case of finite intensity of choice, where most, but not all agents switch to the best performing rule. Because the system is linear in fractions it is of interest to look first at the other limiting case where the intensity of choice is zero. Here always half of the agents are naive, and half are fundamentalistic about each variable. Proposition 10 describes the dynamics of this system. Its proof is given in Appendix C.5.

Proposition 10. *Assume $b = 0$. If $\kappa \leq \frac{\sigma}{2}$, the system described by* (28), (29), (18) *and* (19) *has a unique, stable steady state with inflation and output gap above their targets. If $\kappa > \frac{\sigma}{2}$ the system has an unstable steady state with strictly negative inflation and output gap. Furthermore, the stable eigenvector of the system then has the same slope as that of the system with $b = +\infty$ and all naive expectations.*

From Proposition 10 it follows that with $b = 0$ all initial conditions lead to recovery when $\kappa \leq \frac{\sigma}{2}$. If $\kappa > \frac{\sigma}{2}$ initial conditions above the stable eigenvector through the steady state lead to recovery while initial conditions below this eigenvector lead to divergence. As stated in the proposition, this eigenvector has the same slope as the thin sloped line in Figure 2, but lies strictly

lower in the (π, x)-plane (since it goes through a steady state with negative inflation and output gap). When the naive heuristic is best performing the set of initial conditions that lead to recovery is therefore strictly larger when $b = 0$ (where half of the agents remain fundamentalists) than when $b = +\infty$.

For finite intensity of choice we must distinguish between two cases: periods where naive agents always perform best, and periods were fundamentalistic agents sometimes perform best. In line with Proposition 8, the naive heuristic must necessary be best performing for a deflationary spiral to arise. When this is the case, the system with finite intensity of choice is a convex combination of the systems with $b = 0$ and $b = +\infty$. It follows that a lower intensity of choice leads to a larger region of initial inflation and output gap from which recovery occurs, and that the sloped dashed line in Figure 2 that separates the deflationary spiral region from the recovery region is moved to the left. The intuition is that a lower intensity of choice results in a significant fraction of fundamentalists (higher credibility), even when inflation and output gap are low for a few periods. These fundamentalists put upwards pressure on output gap and inflation, and thereby prevent divergence for initial conditions where a deflationary spiral would have occurred for infinite intensity of choice. However, as the deflationary spiral continues, more an more agents become naive so that eventually (almost) all agents are naive just as in Section 4.1.

If, however, at some point in time fundamentalistic expectations perform better than naive expectations, a lower intensity of choice leads to less fundamentalists. It may then be that for some initial conditions recovery is assured in the infinite intensity of choice case, but divergence occurs for finite intensity of choice. To be more precise, most results for recovery of the previous section hinge on all fundamentalistic expectations. For finite intensity of choice it never happens that all agents become fundamentalistic. Aggregate expectations could therefore be negative even when most agents are fundamentalistic. Recovery now no longer trivially occurs when fundamentalism is the best performing heuristic. Instead, additional constraints on inflation and output gap not being too low are needed to ensure recovery for finite intensity of choice. When most agents follow the central bank, liquidity traps are therefore less likely to occur, but they are still possible.

5 Monetary policy and liquidity traps

Shocks to inflation and output gap can push the economy into a liquidity trap by triggering low self-fulfilling expectations. How can monetary policy prevent these self-fulfilling liquidity traps? In this section we address this question with stochastic simulations. These simulations serve two purposes. First, they illustrate in an intuitive way how stochastic shocks can push our economy with heterogeneous expectations and a zero lower bound on the interest rate into an expectation driven liquidity trap (Section 5.2). Secondly, we study the effectiveness of an increased inflation target, aggressive monetary easing, and aggressive inflation targeting in preventing liquidity traps (Section 5.3).

5.1 Calibration

Unless stated otherwise, the following calibration is used. For κ and σ we use the Woodford (1999) calibration with $\kappa = 0.024$ and $\sigma = 0.157$, and we set $\beta = 0.99$. We further use optimal policy (as defined by (7) and (8)). In this policy rule we set the inflation target equal to an annualized 2%, and, following McCallum and Nelson (2004) and Walsh (2003), the weight on output gap is given by $\mu = 0.25$. The optimal monetary policy coefficients therefore are as given in the first row of Table 1: $\phi_1 = 1.015$ and $\phi_2 = 0.157$.

The shocks to inflation (e_t) and to output gap (u_t), presented in Equations (1) and (2) are reintroduced in this section. Both e_t and u_t are defined as Gaussian white noise and are calibrated to have an annualized standard deviation of 0.01. With this calibration liquidity traps do arise, but they are not so severe that no reasonable police measure can prevent them. The same random seed will be used throughout this section.

Finally, the parameters of the heuristic switching model need to be calibrated. We set the memory parameter in the fitness measure, (12), to $\rho = 0.5$, allowing agents to update their evaluation of the heuristics significantly when new information arises, but also to put considerable weight on the past.[7] The intensity of choice is set to $b = 40.000$, so that it is possible that almost

[7]We also ran all simulations in this section with $\rho = 0$. This only changes result quantitatively. The policy measures presented in Section 5.3 still work to prevent liquidity traps with a lower memory parameter, but the magnitude of the policy change needed to achieve this is larger in that case.

all agents switch to the same heuristic, but that typically both fundamentalists and naive agents will be present.[8]

5.2 The effect of the zero lower bound

Because of the presence of shocks in the model, inflation and output gap no longer exactly converge to a steady state, but fluctuate around it. First we simulate the model for 100 periods, assuming there is no zero lower bound on the nominal interest rate. In Figure 3 the time series of annualized inflation (upper left panel, blue curve) and annualized output gap (upper right panel) are plotted, together with the fractions of fundamentalists (Credibility) for both inflation (middle left panel) and output gap (middle right panel). The bottom panel depicts the annualized nominal interest rate, and the horizontal green line in the upper left panel indicates the annualized inflation target.

Figure 3 illustrates that there are periods where inflation fluctuates around the target, and periods where inflation drifts away. The intuition behind these drifts is the following. When shocks lead to inflation below target for a few consecutive periods, the central bank loses credibility and most agents become naive with respect to inflation (as can be seen in the middle left panel of Figure 3, where the fraction of fundamentalists about inflation moves towards 0). The low expectations of naive agents put downward pressure on inflation and become self-fulfilling. Meanwhile, the central bank tries to control inflation by decreasing the interest rate, but does not immediately succeed. This is so because the CB also cares about output gap, and does not want to react too strongly to inflation expectations in order to limit variations in the output gap. Indeed, we see that output gap stays very close to zero during all periods, and eventually inflation returns to its target as well.

Note that during the downward drifts of inflation the central bank sets a negative interest rate, which in practice cannot happen. In Figure 4 the zero lower bound on the nominal interest rate is accounted for. Now the interest rate is set to zero when it would otherwise have been set negative. All variables evolve in exactly the same way as in Figure 3 until the point where the interest rate

[8]Note that the calibration of the intensity of choice depends on the unit of measurement of the fitness measure. Since a 1% deviation of inflation from steady state is measured as 0.01, and results in a squared forecast error of 0.0001, an intensity of choice of 40.000 should not be considered particularly large.

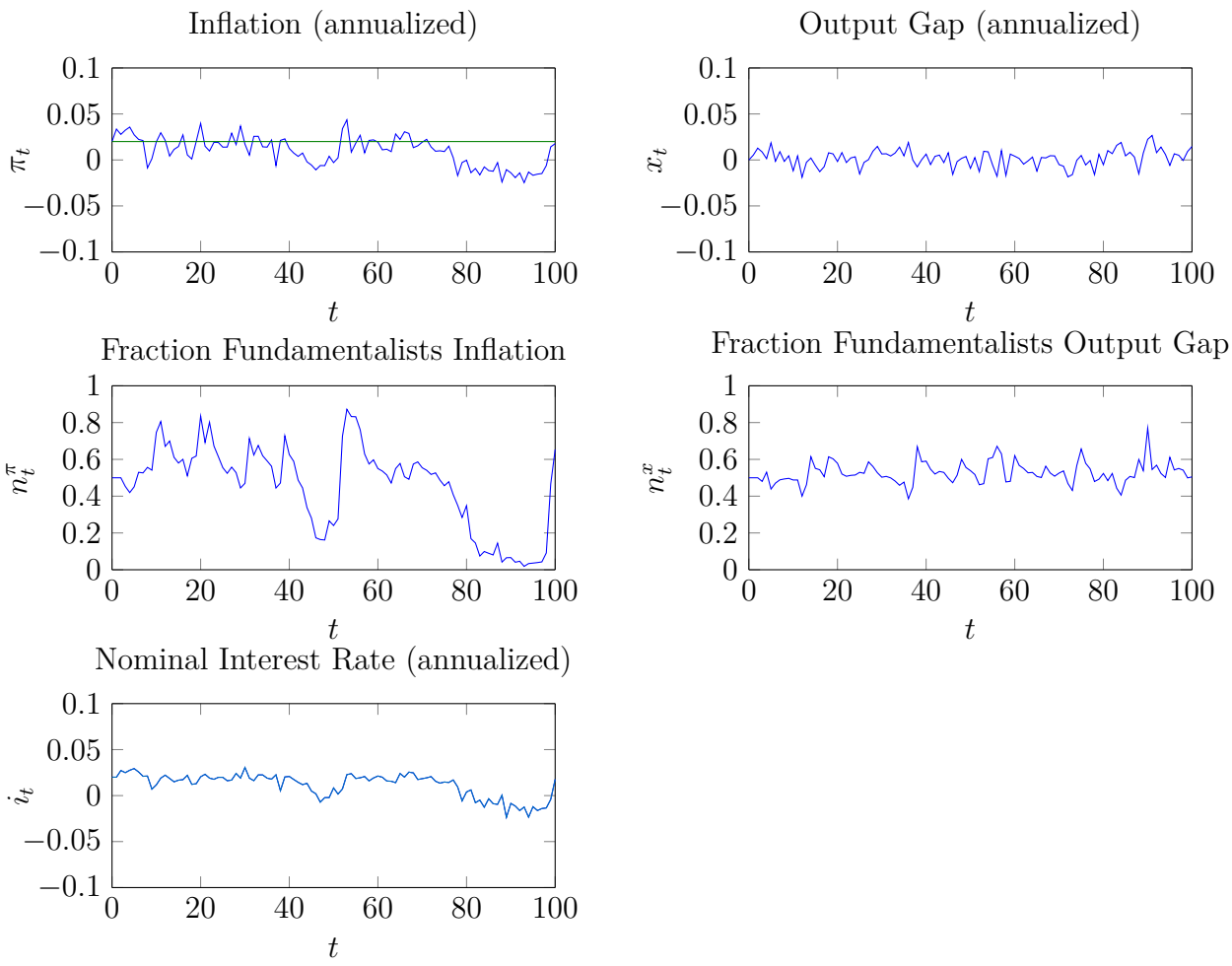

Figure 3: Simulated time series of model with optimal policy, with no lower bound on the nominal interest rate. The horizontal green line in the upper left panel depicts the inflation target of 2%.

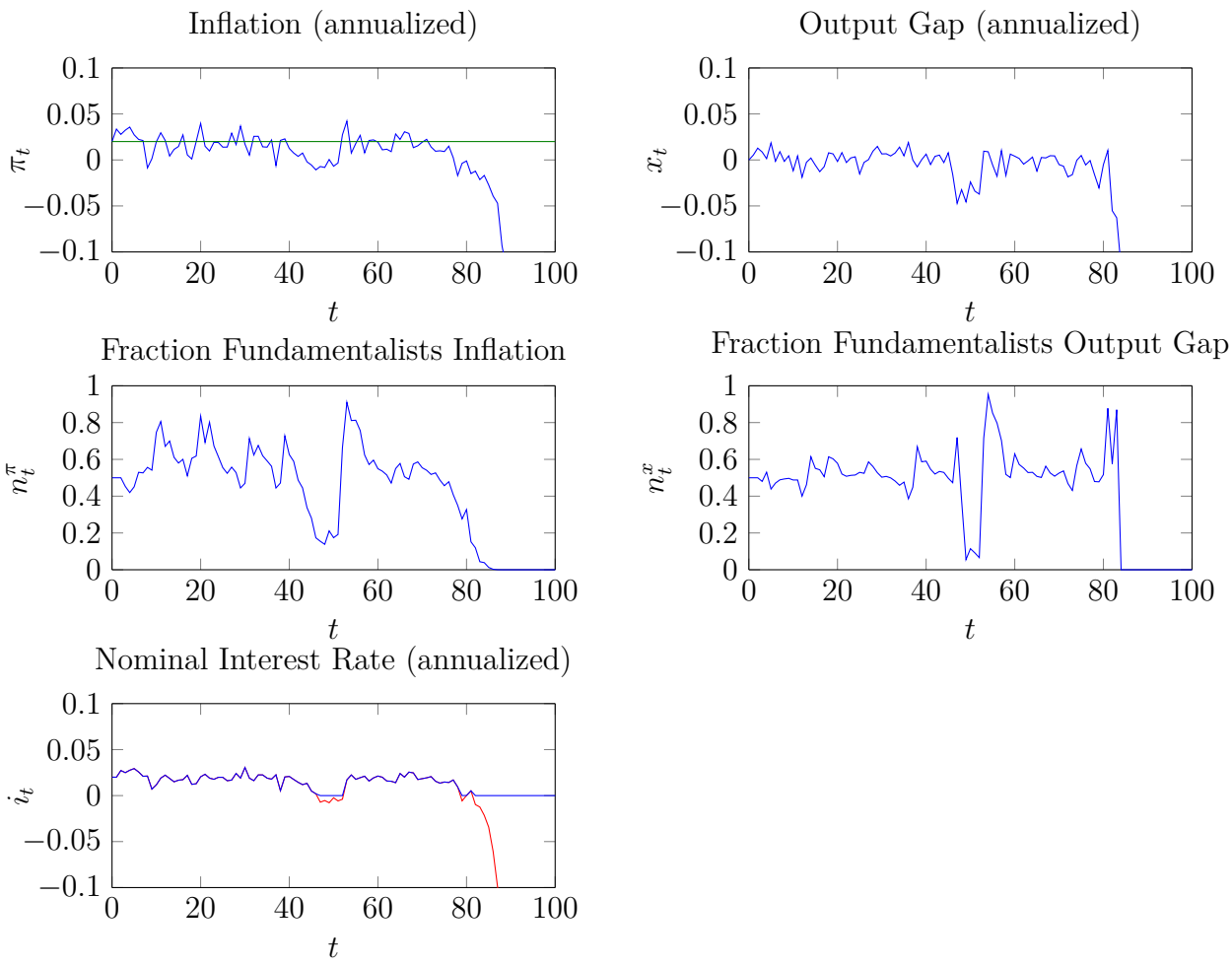

Figure 4: Simulated time series of model with optimal policy, and with ZLB on nominal interest rate. The horizontal green line in the upper left panel depicts the inflation target of 2%. The bottom panel depicts the actual interest rate (blue) and the rate prescribed by (7) (red).

should be set negative. The blue curve in the bottom panel again depicts the annualized nominal interest rate. The red curve depicts the value the interest rate would have taken if it could be set negative. When the ZLB is binding the interest rate is set higher than optimal. In the first liquidity trap around period 45 the low inflation expectations and a higher than optimal nominal interest rate imply a low real interest rate, which depresses output gap. Therefore, in contrast with Figure 3, the economy now enters a recession. Furthermore, the CB now loses credibility with respect to output gap as well as inflation, implying that the economy is in the deflationary spiral case of Table 2 in Section 4.1. However, because of the finite intensity of choice, the CB has not lost all its credibility. This, together with the fact that inflation and output gap are not too low implies that the economy is in the recovery region of Figure 2 and eventually moves back to the fundamental steady state.

In contrast, in the second liquidity from period 78 onwards, inflation, output and credibility decline too much. The economy now is in the deflationary spiral region of Figure 2, and inflation and output gap subsequently diverge towards minus infinity: the system has entered a self-fulfilling deflationary spiral.

5.3 Preventing deflationary spirals

How can the central bank prevent such a deflationary spiral? One possible solution would be to respond with aggressive monetary easing as soon as a liquidity trap is imminent. The central bank could set the interest rate as low as possible (i.e. zero) as soon as it would otherwise have set the interest rate below some threshold. This threshold indicates a danger zone of a low interest rates that threaten to fall below zero.

It turns out that this type of aggressive monetary policy can prevent deflationary spirals in our model, but that the threshold must be chosen high enough in order to be effective. However, if the threshold is set too close to the inflation target (the long run average of the nominal interest rate) the interest rate is set to zero too often, which leads to undesirable output gap fluctuations. The central bank should therefore complement its aggressive monetary easing policy with a higher inflation target. In addition to facilitating this type of policy, an increased inflation target also

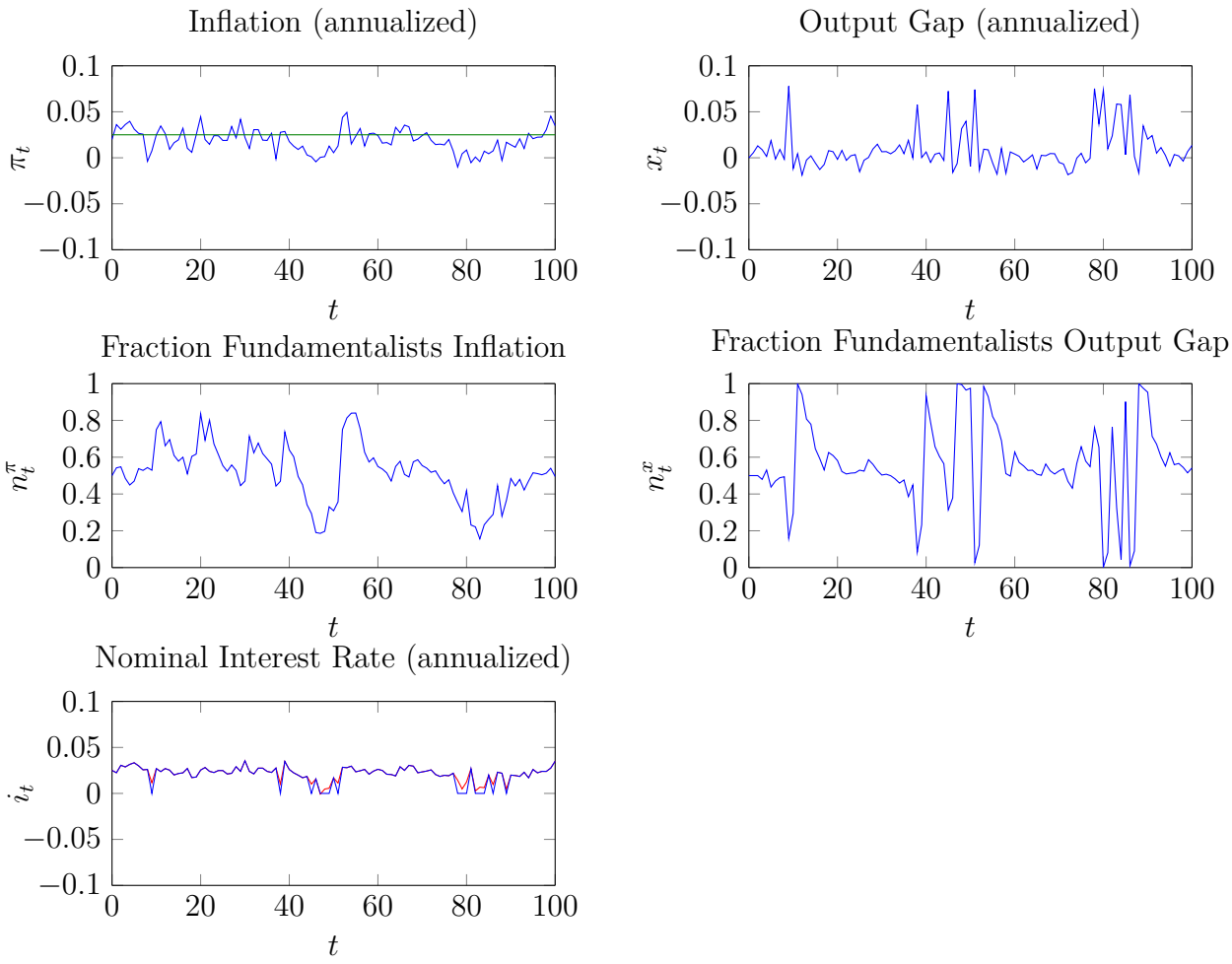

Figure 5: Simulated time series of model with increased inflation target of 2.5% (green) and aggressive monetary easing: the interest rate is set to 0 when it falls below the threshold of 1.5%. The bottom panel depicts the actual interest rate (blue) and the rate prescribed by (7) (red).

makes it less likely that shocks or inflation drifts cause the zero lower bound to be binding, and thereby reduces the frequency of liquidity traps.

Figure 5 shows that a combination of an increased inflation target and aggressive monetary easing indeed can prevent deflationary spirals. The inflation target is here set to an annualized 2.5%, and the central bank conducts aggressive monetary easing as soon as the annualized interest rate would have fallen below 1.5%. In the second episode of low inflation (from period 78 onward), the higher inflation target and the aggressive monetary easing work together to prevent a deflationary spiral. Here inflation falls so much that even with the increased target a liquidity trap would have arisen without aggressive monetary policy. The zero interest rate during many of the following periods brings inflation closer to its target, and increases expectations enough to bring the system

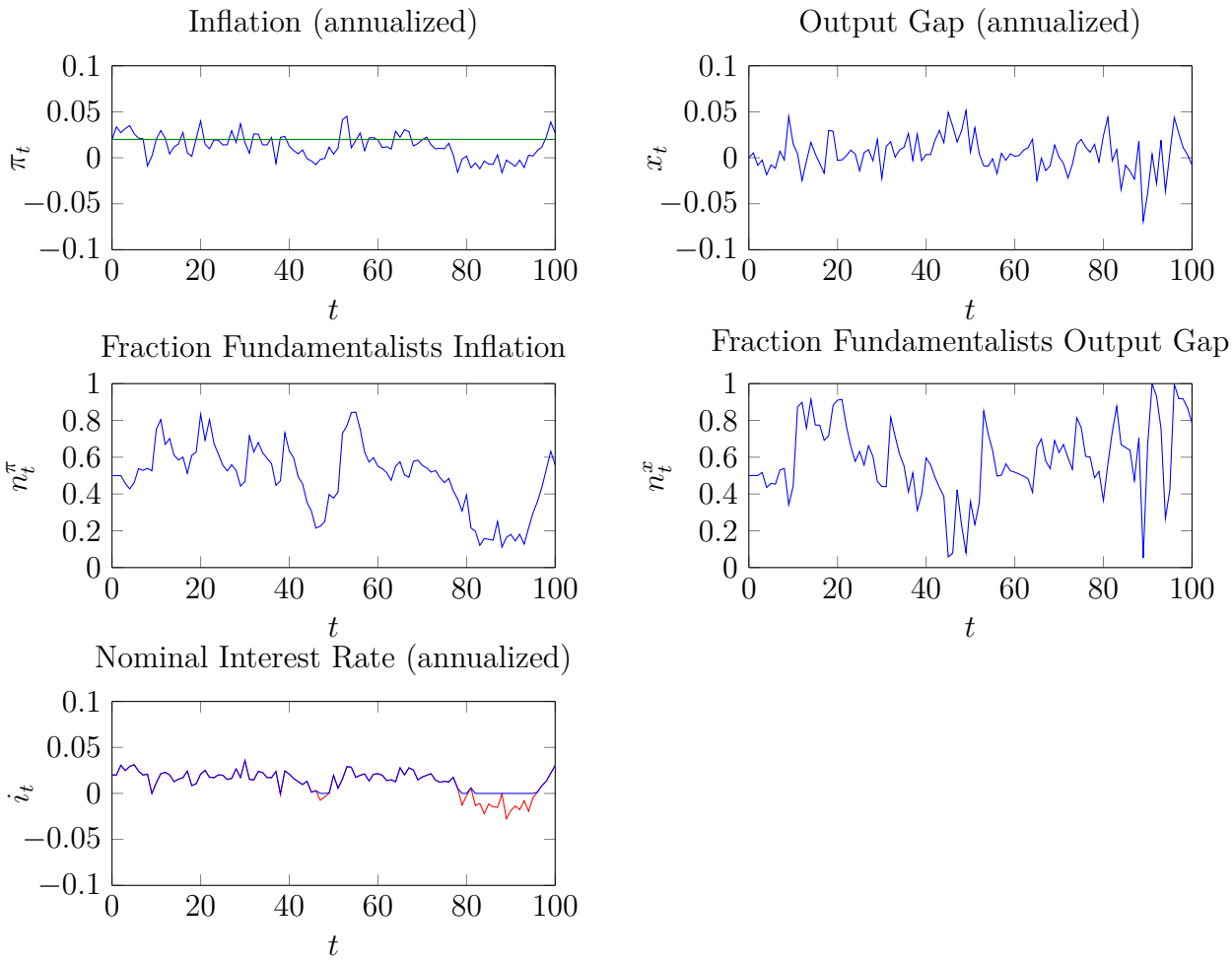

Figure 6: Simulated time series of model with aggressive inflation targeting ($\phi_1 = 1.5$). This corresponds to optimal monetary policy with a very low weight on output gap ($\mu = 0.007$). The horizontal green line in the upper left panel depicts the inflation target of 2%. The bottom panel depicts the actual interest rate (blue) and the rate prescribed by (7) (red).

back to the recovery region. Note, that output gap is severely increased during the periods of monetary easing, but that this is necessary to prevent a deflationary spiral. Alternatively, the inflation target could have been raised even more, but this also comes with its welfare costs.

A final measure that the central bank can take to prevent deflationary spirals is to respond more aggressively to inflation expectations. After all, the liquidity traps arise because inflation is allowed to slowly drift away from its target for several periods under the current policy specification.

Figure 6 plots the case where $\phi_1 = 1.5$. This can either be interpreted as reacting more strongly to inflation than would have been optimal without the zero lower bound, or as optimal policy with a weight on output gap of approximately $\mu = 0.007$. Even though it perhaps is an unrealistic

assumption that the central bank cares over one hundred times more about minimizing inflation deviations than about minimizing output gap deviations, one could argue that, in light of the liquidity trap analysis above, it is much more important to stabilize inflation than it is to stabilize output gap. This point was also made in the analytical analysis in Section 4, where we concluded that inflation expectations play a much large role than output gap expectations in determining whether or not the economy can recover from a liquidity trap.

As a result of the higher inflation coefficient, the interest rate in Figure 6 tracks short term inflation fluctuations much more than in the previous figures. Any negative shock to inflation is immediately countered by a very low interest rate in the next period, which increases inflation again. As a result, drifts in inflation are less severe. The liquidity trap around period 45 has almost disappeared and is accompanied by an increased output gap instead of a recession. The second liquidity trap is still present but inflation, output gap and credibility do not decline as much as in Figure 4, so that the economy stays in the recovery region of Figure 2, and a deflationary spiral does not arise.

Note that the cost of the increased inflation coefficient in the Taylor rule arise in the form of stronger output gap fluctuations (also in times where no liquidity trap is imminent) than in Figure 3, which is consistent with the fact that we are now considering optimal monetary policy with a very low weight on output gap. As in the case of aggressive monetary easing, these increased output gap fluctuations can be seen as a sacrifice, necessary to prevent deflationary spirals.

6 Conclusion

In this paper we use a New Keynesian model to study optimal inflation targeting and liquidity traps. Instead of assuming rational expectations, we allow expectations to be formed heterogeneously by using a model where agents switch between heuristics based on relative performance. In our model, fundamentalists, who thrust the central bank, compete with naive agents, who base their forecast on past information. We therefore can interpret the fraction of fundamentalistic agents as the credibility of the central bank. Unlike in rational expectations models, this allows us to *endogenously* model the central banks' credibility, which is of crucial importance in

understanding liquidity traps.

Our first finding is that a nominal interest rate that responds too weakly or too strongly to output gap expectations leads to instability of the fundamental steady state. In this steady state both inflation and output gap are equal to the targets set by the central bank. The region of policy parameters that lead to local stability of the fundamental steady state is however strictly larger than the region of policy parameters that result in a locally determinate equilibrium when rational expectations are assumed. In fact, we find that the well known Taylor principle is not a necessary condition for local stability in our heterogeneous expectations model.

We furthermore show that the fundamental steady state is always globally stable and unique when a theoretically optimal interest rate rule is used that is derived from a loss function with all future output gaps and all future deviations of inflation from its target. This policy specification is however only optimal when the central bank is not hindered by the zero lower bound on the nominal interest rate.

When the zero lower bound is introduced to our model, we find that expectation driven liquidity traps can occur, even under optimal policy. In such a liquidity trap, the central bank has lost some, or all, of its credibility, and low, naive expectations make the zero lower bound constraint binding. Whether the economy can recover form such a liquidity trap, or whether a deflationary spiral with ever decreasing inflation and output gap occurs, depends critically both on how low inflation and output gap have become, and on how much credibility the central bank is able to retain. If the central bank has lost too much credibility, and inflation and output gap are too low, more and more agents start to coordinate on low, naive expectations, resulting in a self-fulfilling deflationary spiral. Coordination on naive expectations or on some other form of adaptive expectations is an empirically relevant and plausible situation that is encountered e.g. in Assenza et al. (2014) and other laboratory experiments.

In stochastic simulations with optimal monetary policy we find that small shocks to the economy can lead to coordination on low naive expectations, and that this can result in both transient liquidity traps and in deflationary spirals. We furthermore show that a central bank can prevent deflationary spirals by increasing the inflation target, applying aggressive monetary easing when

the interest rate becomes too low, or by responding more aggressively to inflation than optimal without a zero lower bound. All these policy measures come however, with their own disadvantages and costs to the economy. Therefore, a well balanced combination of all three measures may be the best way to proceed.

In future work, we will further study the robustness of our results under different heterogeneous forecasting rules. One approach is to study monetary policy in a large type limit (LTL) macroeconomy, where the number of heterogeneous rules tends to infinity (Brock et al., 2005). We furthermore plan to develop models where the role of *forward guidance* and *fiscal policy* in escaping liquidity traps can be investigated under heterogeneous expectations.

A Microfoundations

The following derivation largely follows the steps of Kurz et al. (2013). We make however use of the properties of our heuristic switching model, which allows us to fully aggregate, and to obtain, under heterogeneous Euler equation learning, the same equations that arise under a representative household with rational expectations.

There is a continuum (i) of households who differ in the way they form expectations about inflation and about output gap. Households with the same expectations have the same preferences and will make the same decisions. The intratemporal problem of each household i, consists of choosing consumption over a continuum of differentiated goods (j) to minimize expenditure. This implies

$$C_t^i(j) = \left(\frac{p_t(j)}{P_t}\right)^{-\theta} C_t^i, \tag{A.1}$$

with C_t^i and P_t total consumption of the household and the aggregate price level, defined by

$$C_t^i = \left(\int_0^1 C_t^i(j)^{\frac{\theta-1}{\theta}} dj\right)^{\frac{\theta}{\theta-1}}, \tag{A.2}$$

$$P_t = \left(\int_0^1 P_t(j)^{1-\theta} dj\right)^{\frac{1}{1-\theta}}, \tag{A.3}$$

where θ is the elasticity of substitution between the different goods.

The household i then chooses consumption (C_t^i), labor (H_t^i), and real bond holdings (b_t^i) to maximize

$$\tilde{E}_t^i \sum_{s=t}^{\infty} \beta^{s-t} \left[\frac{(C_s^i)^{1-\sigma}}{1-\sigma} - \frac{(H_s^i)^{1+\eta}}{1+\eta}\right], \tag{A.4}$$

subject to its budget constraint

$$C_t^i + b_t^i \ \leq w_t H_t^i + \frac{b_{t-1}^i(1+i_{t-1})}{1+\pi_t} + T_t^i, \tag{A.5}$$

where β_t is the discount factor, w_t the real wage rate, i_t the nominal interest rate, $\pi_t = \frac{P_t}{P_{t-1}} - 1$ is the inflation rate, and T_t^i real lump sum transfers to household i, including profits from firms. $\tilde{E}_t^i$ represents the subjective expectation operator that differs over the households.

The first order conditions with respect to C_t^i, H_t^i and b_t^i give

$$(C_t^i)^{-\sigma} = \lambda_t^i,$$

$$(H_t^i)^{\eta} = \lambda_t(1-\tau_H)w_t,$$

$$\lambda_t^i = \beta \tilde{E}_t^i \frac{\lambda_{t+1}^i(1+i_t)}{1+\pi_{t+1}},$$

with λ the Lagrange multiplier. Solving for this multiplier, we can rewrite these conditions to the Euler equation and an expression for the real wage rate, which, together with the budget constraint (A.5), must hold in equilibrium

$$(C_t^i)^{-\sigma} = \beta \tilde{E}_t^i \left[\frac{(C_{t+1}^i)^{-\sigma}(1+i_t)}{1+\pi_{t+1}} \right], \tag{A.6}$$

$$w_t = (H_t^i)^{\eta}(C_t^i)^{\sigma}. \tag{A.7}$$

The Euler equation, (A.6), can be log linearized around a zero inflation steady state to get

$$\hat{C}_t^i = \tilde{E}_t^i[\hat{C}_{t+1}^i] - \frac{1}{\sigma}(i_t - \tilde{E}_t^i[\pi_{t+1}]), \tag{A.8}$$

where $\hat{C}_t = \frac{C_t - \bar{C}}{\bar{C}}$, with $\bar{C}$ the steady state value of consumption.

We assume that our boundedly rational agents use Euler equation learning (see Honkapohja et al., 2012), implying that they use the two period trade-off of (A.8) to make optimal decisions given their subjective forecasts of next period. Microfoundations with heterogeneous expectations under infinite horizon learning are derived by Massaro (2013).

Next, we deviate form Kurz et al. (2013), and use a property of the discrete choice model (Equation (11)), which determines the fractions of agents in each period as in Brock and Hommes (1997). Under this model it is implicitly assumed that the probability to follow a particular heuristic next period is the same across agents, i.e., independent of the heuristic they followed in the past. This reflects the fact that our agents are not inherently different, but that each of them faces the same trade-off between becoming naive or fundamentalist each period. We assume agents know (have learned) that all agents have the same probability to follow a particular heuristic in the future, and that they know that consumption decisions only differ between households in so

far as their expectations are different. In that case households' expectations about their own future consumption coincide with their expectations about the future consumption of any other agent, and therefore with aggregate consumption. That is, $\tilde{E}_t^i[\hat{C}_{t+1}^i] = \tilde{E}_t^i[\hat{C}_{t+1}]$, with $\hat{C}_{t+1} = \int_0^1 \hat{C}_{t+1}^i di$. Agents therefore realize they should base their current period consumption decision on expectations about future *aggregate* consumption. The Euler equation can then be written as

$$\hat{C}_t^i = \tilde{E}_t^i[\hat{C}_{t+1}] - \frac{1}{\sigma}(i_t - \tilde{E}_t^i[\pi_{t+1}]), \tag{A.9}$$

Market clearing in each good j market imposes that

$$Y_t(j) = C_t(j), \tag{A.10}$$

where $C_t(j) = \int C_t^i(j) di$ is aggregate consumption of good j. If we aggregate over all varieties of goods, we end up with the aggregate goods market clearing condition

$$Y_t = C_t. \tag{A.11}$$

We assume that agents have learned about market clearing, so that their forecasts satisfy $\tilde{E}_t^i[\hat{C}_{t+1}] = \tilde{E}_t^i[\hat{Y}_{t+1}]$. Therefore, (A.9) can be written as

$$\hat{C}_t^i = \tilde{E}_t^i[\hat{Y}_{t+1}] - \frac{1}{\sigma}(i_t - \tilde{E}_t^i[\pi_{t+1}]), \tag{A.12}$$

Aggregating this equation over all agents, and using the period t market clearing condition then gives

$$\hat{Y}_t = \bar{E}_t[\hat{Y}_{t+1}] - \frac{1}{\sigma}(i_t - \bar{E}_t[\pi_{t+1}]) \tag{A.13}$$

Here $\bar{E}_t$ is the aggregate expectation operator defined as $\bar{E}_t[Z_{t+1}] = n_t^Z \tilde{E}_t^F[Z_{t+1}] + (1 - n_t^N)\tilde{E}_t^N[Z_{t+1}]$, with $\tilde{E}_t^F$ the fundamentalist expectation operator, $\tilde{E}_t^N$ the naive expectation operator, and n_t^Z the fraction of agents that are fundamentalist with respect to variable Z.

Next we turn to the supply side of the economy. There is a continuum (j) of firms producing the differentiated goods. Each firm is run by a household and follows the same heuristics for prediction of future variables as that household in each period. Each firm has a linear technology

with labor as its only input

$$Y_t(j) = A_t H_t(j), \tag{A.14}$$

where A_t is aggregate productivity in period t. We assume that in each period a fraction $(1-\omega)$ firms can change their price, as in Calvo (1983). Firms want to choose the price $p(j)$ that maximizes their expected discounted profits

$$\tilde{E}_t^j \sum_{s=0}^{\infty} \omega^s Q_{t,t+s}^j \left[p_t(j) Y_{t+s}(j) - P_{t+s} mc_{t+s} Y_{t+s}(j) \right], \tag{A.15}$$

where

$$Q_{t,t+s}^j = \beta^s \left(\frac{C_{t+s}^j}{C_t^j} \right)^{-\sigma} \frac{P_t}{P_{t+s}}. \tag{A.16}$$

is the stochastic discount factor of the household (j) that runs firm j.

$$mc_t = \frac{w_t(1-\nu)}{A_t}, \tag{A.17}$$

are real marginal cost incurred by firms, with ν a production subsidy. Using the demand for good j, the firm's profits maximization problem writes as follows

$$\max \tilde{E}_t^j \sum_{s=0}^{\infty} \omega^s \beta^s \left(\frac{C_{t+s}^j}{C_t^j} \right)^{-\sigma} P_t \left[\left(\frac{p_t(j)}{P_{t+s}} \right)^{1-\theta} Y_{t+s} - mc_{t+s} \left(\frac{p_t(j)}{P_{t+s}} \right)^{-\theta} Y_{t+s} \right]. \tag{A.18}$$

The first order condition for $p_t(j)$ is

$$\tilde{E}_t^j \sum_{s=0}^{\infty} \omega^s \beta^s \left(\frac{C_{t+s}^j}{C_t^j} \right)^{-\sigma} \frac{P_t}{P_{t+s}} \left(\frac{p_t^*(j)}{P_{t+s}} \right)^{-1-\theta} Y_{t+s} \left[\frac{p_t^*(j)}{P_{t+s}} - \frac{\theta}{\theta - 1} mc_{t+s} \right] = 0, \tag{A.19}$$

where $p_t^*(j)$ is the optimal price for firm j if it can re-optimize in period t.

This can be written as

$$q_t^*(j) \tilde{E}_t^j \sum_{s=0}^{\infty} \omega^s \beta^s \left(C_{t+s}^j \right)^{-\sigma} \left(\frac{P_{t+s}}{P_t} \right)^{\theta-1} Y_{t+s} = \frac{\theta}{\theta - 1} \tilde{E}_t^j \sum_{s=0}^{\infty} \omega^s \beta^s \left(C_{t+s}^j \right)^{-\sigma} \left(\frac{P_{t+s}}{P_t} \right)^{\theta} Y_{t+s} mc_{t+s}, \tag{A.20}$$

with $q_t^*(j) = \frac{p_t^*(j)}{P_t}$.

Log linearizing gives

$$\frac{\hat{q}_t^*(j)}{1-\omega\beta} = \tilde{E}_t^j \sum_{s=0}^{\infty} \omega^s \beta^s (\hat{mc}_{t+s} + \hat{p}_{t+s}) - \frac{1}{1-\omega\beta}\hat{p}_t, \tag{A.21}$$

which can be written as

$$\hat{q}_t^*(j) + \hat{p}_t = (1-\omega\beta)(\hat{mc}_t + \hat{p}_t) + \omega\beta(1-\omega\beta)\tilde{E}_t^j \sum_{s=0}^{\infty} \omega^s \beta^s (\hat{mc}_{t+s+1} + \hat{p}_{t+s+1}), \tag{A.22}$$

or recursively as

$$\begin{aligned} \hat{q}_t^*(j) + \hat{p}_t &= (1-\omega\beta)(\hat{mc}_t + \hat{p}_t) + \omega\beta\tilde{E}_t^j[\hat{q}_{t+1}^*(j) + \hat{p}_{t+1}], \\ \hat{q}_t^*(j) &= (1-\omega\beta)\hat{mc}_t + \omega\beta\tilde{E}_t^j[\hat{q}_{t+1}^*(j) + \pi_{t+1}]. \end{aligned} \tag{A.23}$$

Just as in the case of consumption, it follows from the discrete choice model that $\tilde{E}_t^j[\hat{q}_{t+1}^*(j)] = \tilde{E}_t^j[\hat{q}_{t+1}^*]$. Therefore, agents base their pricing decisions on their expectations of future aggregate variables, and we can write

$$\hat{q}_t^*(j) = (1-\omega\beta)\hat{mc}_t + \omega\beta\tilde{E}_t^j[\hat{q}_{t+1}^* + \pi_{t+1}]. \tag{A.24}$$

Next we turn to the evolution of the aggregate price level. We assume that the set of firms that can change their price in a period is chosen independently of the types of the households running the firm, so that the distribution of expectations of firms that can change their price is identical to the distribution of expectations of all firms. Since decisions of firms only differ in so far their expectations differ, it follows that the aggregate price level evolves as

$$P_t = [\omega P_{t-1}^{1-\theta} + (1-\omega)\int_0^1 p_t^*(j)^{1-\theta} dj]^{\frac{1}{1-\theta}}, \tag{A.25}$$

This can be log linearized to

$$\hat{p}_t = \omega\hat{p}_{t-1} + (1-\omega)\int_0^1 \hat{p}_t^*(j) dj, \tag{A.26}$$

from which it follows that

$$\frac{\omega}{1-\omega}\pi_t = \int_0^1 \hat{q}_t^*(j) dj = \hat{q}_t^*. \tag{A.27}$$

Plugging this into (A.24) gives

$$\hat{q}_t^*(j) = (1-\omega\beta)\hat{mc}_t + \frac{\omega\beta}{1-\omega}\tilde{E}_t^j[\pi_{t+1}]. \tag{A.28}$$

Aggregating over all firms and again using (A.27) gives

$$\pi_t = \beta\bar{E}_t[\pi_{t+1}] + \kappa\hat{mc}_t, \tag{A.29}$$

with

$$\tilde{\kappa} = \frac{(1-\omega)(1-\beta\omega)}{\omega} \tag{A.30}$$

Log linearizing (A.17), (A.7) and (A.14), and combining with market clearing gives

$$\hat{mc}_t = \hat{w}_t - \hat{A}_t = \eta\hat{H}_t + \sigma\hat{C}_t - \hat{A}_t = (\sigma+\eta)\hat{Y}_t - (1+\eta)\hat{A}_t \tag{A.31}$$

Inserting this in (A.29) results in

$$\pi_t = \beta\bar{E}_t[\pi_{t+1}] + \tilde{\kappa}(\sigma+\eta)\hat{Y}_t - \tilde{\kappa}(1+\eta)\hat{A}_t, \tag{A.32}$$

Finally we write (A.13) and (A.32) in terms of output gap. Here we assume that the subsidy to firms offsets the distortions due to monopolistic competition, so that the flexible price equilibrium is efficient.

It follows from (A.31) that the potential level of output is given by

$$\hat{Y}_t^{pot} = \frac{(1+\eta)}{\sigma+\eta}\hat{A}_t \tag{A.33}$$

Plugging in $x_t = \hat{Y}_t - \hat{Y}_t^{pot}$ in (A.13) and (A.32) gives

$$x_t = \bar{E}_t[x_{t+1}] - \frac{1}{\sigma}(i_t - \bar{E}_t[\pi_{t+1}]) + u_t \tag{A.34}$$

$$\pi_t = \beta\bar{E}_t[\pi_{t+1}] + \kappa x_t, \tag{A.35}$$

with $\kappa = \tilde{\kappa}(\sigma+\eta)$, and $u_t = \frac{(1+\eta)}{\sigma+\eta}(\hat{A}_{t+1} - \hat{A}_t)$. If we introduce a cost push shock (e_t) in the Phillips curve the model of (1) and (2) is obtained.

B Monetary policy without the ZLB

B.1 Proof Proposition 1

Equation (16) and (17) can be combined to give the following equation that holds in steady state.

$$\left(1-\beta\frac{(1-m^{\pi})}{2}+\kappa\frac{\frac{\phi_1-1}{\sigma}\frac{(1-m^{\pi})}{2}}{(1-(1-\frac{\phi_2}{\sigma})\frac{(1-m^{x})}{2})}\right)\pi=\left(\beta\frac{(1+m^{\pi})}{2}+\kappa\frac{\frac{\phi_1-1}{\sigma}\frac{(1-m^{\pi})}{2}}{(1-(1-\frac{\phi_2}{\sigma})\frac{(1-m^{x})}{2})}\right)\pi^{T}, \tag{B.1}$$

with

$$m^{x}=tanh(-\frac{b}{2}x^{2}) \tag{B.2}$$

$$m^{\pi}=tanh(-\frac{b}{2}(\pi-\pi^{T})^{2}) \tag{B.3}$$

For $\pi^T = 0$, or for $\beta \to 1$ this has as a solution the fundamental steady state: $x^* = 0, \pi^* = \pi^T, m^{x*} = 0, m^{\pi*} = 0$. For $\pi^T > 0$ and β close to 1, the solution lies very close to the above values. More specifically, we then have:

$$\pi^{*}=\frac{\beta\frac{(1+m^{\pi})}{2}+\kappa\frac{\frac{\phi_1-1}{\sigma}\frac{(1-m^{\pi})}{2}}{(1-(1-\frac{\phi_2}{\sigma})\frac{(1-m^{x})}{2})}}{1-\beta\frac{(1-m^{\pi})}{2}+\kappa\frac{\frac{\phi_1-1}{\sigma}\frac{(1-m^{\pi})}{2}}{(1-(1-\frac{\phi_2}{\sigma})\frac{(1-m^{x})}{2})}}\pi^{T}.$$

B.2 Jacobian and eigenvalues

In this section, first the Jacobian of the system given by by (16) through (19) is presented. Next this Jacobian is evaluated at the fundamental steady state, and eigenvectors are derived.

The Jacobian is given by

$$\begin{pmatrix}
(1-\frac{\phi_2}{\sigma})\frac{(1-m_t^{x})}{2} & -\frac{\phi_1-1}{\sigma}\frac{(1-m_t^{\pi})}{2} & 0 & 0 & -\frac{1}{2}(1-\frac{\phi_2}{\sigma})x_{t-1} & \frac{\phi_1-1}{2\sigma}\left(\pi_{t-1}-\pi^{T}\right)\\
\kappa(1-\frac{\phi_2}{\sigma})\frac{(1-m_t^{x})}{2} & \beta\frac{(1-m_t^{\pi})}{2}-\kappa\frac{\phi_1-1}{\sigma}\frac{(1-m_t^{\pi})}{2} & 0 & 0 & -\frac{\kappa}{2}(1-\frac{\phi_2}{\sigma})x_{t-1} & (\kappa\frac{\phi_1-1}{2\sigma}-\frac{\beta}{2})\left(\pi_{t-1}-\pi^{T}\right)\\
1 & 0 & 0 & 0 & 0 & 0\\
0 & 1 & 0 & 0 & 0 & 0\\
c_{11}s_A & c_{12}s_A & d_{11}s_A & 0 & e_{11}s_A & e_{12}s_A\\
c_{21}s_B & c_{22}s_B & 0 & d_{22}s_B & e_{21}s_B & e_{22}s_B
\end{pmatrix}$$

with

$$C = \begin{pmatrix} -(1-\frac{\phi_2}{\sigma})(1-m_t^x)x_{t-2} & \frac{\phi_1-1}{\sigma}(1-m_t^\pi)x_{t-2} \\ -\kappa(1-\frac{\phi_2}{\sigma})(1-m_t^x)\left(\pi_{t-2}-\pi^T\right) & (\kappa\frac{\phi_1-1}{\sigma}-\beta)(1-m_t^\pi)\left(\pi_{t-2}-\pi^T\right) \end{pmatrix}$$

$$E = \begin{pmatrix} (1-\frac{\phi_2}{\sigma})x_{t-1}x_{t-2} & -\frac{\phi_1-1}{\sigma}\left(\pi_{t-1}-\pi^T\right)x_{t-2} \\ \kappa(1-\frac{\phi_2}{\sigma})x_{t-1}\left(\pi_{t-2}-\pi^T\right) & -(\kappa\frac{\phi_1-1}{\sigma}-\beta)\left(\pi_{t-1}-\pi^T\right)\left(\pi_{t-2}-\pi^T\right) \end{pmatrix}$$

$$d_{11} = 2x_{t-2} - 2x_t$$

$$d_{22} = 2\pi_{t-2} - 2\pi_t$$

$$s_A = \frac{b}{2}sech\left(\frac{b}{2}(x_{t-2}^2 - 2x_t x_{t-2})\right)$$

$$s_B = \frac{b}{2}sech\left(\frac{b}{2}(\pi_{t-2}^2 - (\pi^T)^2 - 2(\pi_{t-2}-\pi^T)\pi_t)\right)$$

with x_t and π_t given by (16) and (17)

In the fundamental steady state where $x_t = 0$, $\pi_t = \pi^T$ and $m_t^x = m_t^\pi = 0$ for all t (either because $\pi^T = 0$, or because $\beta \to 1$), the Jacobian reduces to

$$\begin{pmatrix} \frac{1}{2}(1-\frac{\phi_2}{\sigma}) & -\frac{1}{2}\frac{\phi_1-1}{\sigma} & 0 & 0 & 0 & 0 \\ \frac{\kappa}{2}(1-\frac{\phi_2}{\sigma}) & \frac{\beta}{2}-\frac{\kappa}{2}\frac{\phi_1-1}{\sigma} & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

This is a lower triangular block matrix, with four eigenvalues equal to 0. The other eigenvalues are the eigenvalues of the upper left 2x2 block. These two eigenvalues are equal to

$$\lambda_1 = \frac{1}{4}\left((1+\beta-\frac{\phi_2}{\sigma}-\kappa\frac{\phi_1-1}{\sigma}) + \sqrt{\left(1+\beta-\frac{\phi_2}{\sigma}-\kappa\frac{\phi_1-1}{\sigma}\right)^2 - 4\beta(1-\frac{\phi_2}{\sigma})}\right) \qquad \text{(B.4)}$$

and

$$\lambda_2 = \frac{1}{4}\left((1+\beta-\frac{\phi_2}{\sigma}-\kappa\frac{\phi_1-1}{\sigma}) - \sqrt{\left(1+\beta-\frac{\phi_2}{\sigma}-\kappa\frac{\phi_1-1}{\sigma}\right)^2 - 4\beta(1-\frac{\phi_2}{\sigma})}\right) \tag{B.5}$$

B.3 Proof Proposition 2

From (B.4) we know that $\lambda_1 > 1$ if and only if

$$\frac{1}{4}\left((1+\beta-\frac{\phi_2}{\sigma}-\kappa\frac{\phi_1-1}{\sigma}) + \sqrt{\left(1+\beta-\frac{\phi_2}{\sigma}-\kappa\frac{\phi_1-1}{\sigma}\right)^2 - 4\beta(1-\frac{\phi_2}{\sigma})}\right) > 1 \tag{B.6}$$

$$\left(1+\beta-\frac{\phi_2}{\sigma}-\kappa\frac{\phi_1-1}{\sigma}\right)^2 - 4\beta(1-\frac{\phi_2}{\sigma}) > 16 + (1+\beta-\frac{\phi_2}{\sigma}-\kappa\frac{\phi_1-1}{\sigma})^2 - 8(1+\beta-\frac{\phi_2}{\sigma}-\kappa\frac{\phi_1-1}{\sigma})$$

$$-(2-\beta)(1+\frac{\phi_2}{\sigma}) > 2\kappa\frac{\phi_1-1}{\sigma}$$

$$\phi_1 < \phi_1^{PF} = 1-(2-\beta)\frac{\phi_2+\sigma}{2\kappa} \tag{B.7}$$

Below we show that a pitchfork bifurcation occurs at this value of ϕ_1 by showing that two non-fundamental symmetric steady states are created here.

Non-fundamental steady states could exist as solutions of (B.1) if they satisfy

$$1-\beta\frac{(1-m^\pi)}{2} + \kappa\frac{\frac{\phi_1-1}{\sigma}\frac{(1-m^\pi)}{2}}{(1-(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2})} = 0 \tag{B.8}$$

$$m^\pi(-\beta+\kappa\frac{\frac{\phi_1-1}{\sigma}}{(1-(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2})}) = 2-\beta+\kappa\frac{\frac{\phi_1-1}{\sigma}}{(1-(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2})} \tag{B.9}$$

$$m^\pi = \frac{(2-\beta)(1-(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2}) + \kappa\frac{\phi_1-1}{\sigma}}{-\beta(1-(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2}) + \kappa\frac{\phi_1-1}{\sigma}} \tag{B.10}$$

The steady state values of π then are

$$\pi^* = \pi^T \pm \sqrt{-\frac{2}{b}tanh^{-1}\left(\frac{(2-\beta)(1-(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2}) + \kappa\frac{\phi_1-1}{\sigma}}{-\beta(1-(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2}) + \kappa\frac{\phi_1-1}{\sigma}}\right)} \tag{B.11}$$

When non-fundamental steady states exist, there thus are two non-fundamental steady states, symmetric around the fundamental value π^T.

Because in a non-fundamental steady states naive predictors perform better than fundamentalists, non-fundamental steady states can only exist with

$$-1 \leq m^{\pi} < 0 \tag{B.12}$$

Since it is assumed that both σ and ϕ_1 are non-negative we must have

$$(1 - (1 - \frac{\phi_2}{\sigma})\frac{(1 - m^x)}{2}) > 0 \tag{B.13}$$

Using this and (B.10), the inequalities in (B.12) reduce to

$$\beta(1 - (1 - \frac{\phi_2}{\sigma})\frac{(1 - m^x)}{2}) - \kappa\frac{\phi_1 - 1}{\sigma} \geq (2 - \beta)(1 - (1 - \frac{\phi_2}{\sigma})\frac{(1 - m^x)}{2}) + \kappa\frac{\phi_1 - 1}{\sigma} > 0 \tag{B.14}$$

or equivalently

$$1 - \frac{\sigma}{\kappa}(1 - \beta)\left(1 - (1 - \frac{\phi_2}{\sigma})\frac{(1 - m^x)}{2}\right) \geq \phi_1 > 1 - \frac{\sigma}{\kappa}(2 - \beta)\left(1 - (1 - \frac{\phi_2}{\sigma})\frac{(1 - m^x)}{2}\right) \tag{B.15}$$

From the equivalence of (B.12) and (B.15), it can be concluded that as ϕ_1 gets close to its right-hand limit, m^{π} gets close to zero. This implies that m^x , x and π also go to their fundamental values as this happens. Using that m^x goes to zero in the limit, we see from Equation (B.15) that the limiting value of ϕ_1 for which the non-fundamental steady state exist is

$$\phi_1^{PF} = 1 - \frac{\sigma}{\kappa}(2 - \beta)(1 - (1 - \frac{\phi_2}{\sigma})\frac{1}{2}) = 1 - (2 - \beta)\frac{\phi_2 + \sigma}{2\kappa} \tag{B.16}$$

At this point both steady states coincide with the fundamental steady state. We can conclude that at the bifurcation value indeed two non-fundamental steady states are created, which exists for values of ϕ_1 larger than ϕ_1^{PF}. The bifurcation therefore is a subcritical pitchfork bifurcation and the non-fundamental steady states must be unstable.

B.4 Proof Proposition 3

λ_2 is smaller than -1 when

$$\frac{1}{4}\left((1+\beta-\frac{\phi_2}{\sigma}-\kappa\frac{\phi_1-1}{\sigma})-\sqrt{\left(1+\beta-\frac{\phi_2}{\sigma}-\kappa\frac{\phi_1-1}{\sigma}\right)^2-4\beta(1-\frac{\phi_2}{\sigma})}\right)<-1 \tag{B.17}$$

Rewriting this as

$$\sqrt{\left(1+\beta-\frac{\phi_2}{\sigma}-\kappa\frac{\phi_1-1}{\sigma}\right)^2-4\beta(1-\frac{\phi_2}{\sigma})}>4+(1+\beta-\frac{\phi_2}{\sigma}-\kappa\frac{\phi_1-1}{\sigma}) \tag{B.18}$$

$$-4\beta(1-\frac{\phi_2}{\sigma})>16+8(1+\beta-\frac{\phi_2}{\sigma}-\kappa\frac{\phi_1-1}{\sigma}), \tag{B.19}$$

the condition reduces to

$$(2+\beta)\frac{\phi_2}{\sigma}>3(2+\beta)-2\kappa\frac{\phi_1-1}{\sigma}, \tag{B.20}$$

which can be rewritten as

$$\phi_1>1+(2+\beta)\frac{3\sigma-\phi_2}{2\kappa}, \tag{B.21}$$

When one eigenvalue becomes -1, a 2-cycle must exists either below or above the bifurcation value. This makes the period doubling bifurcation either subcritical or supercritical. In what follows ϕ_1 is treated as the bifurcation parameter. The value of ϕ_2 then turns out to determine if the bifurcation is subcritical or supercritical.

The 2-cycle in question is symmetric around the fundamental steady state. We thus have $x_1=-x_2$ and $(\pi_1-\pi^T)=-(\pi_2-\pi^T)$ (where we again assume that either $\pi^T=0$, or $\beta\to 1$). Using this, (5) and (17) can be written as

$$x=x_1=(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2}x_2-\frac{\phi_1-1}{\sigma}\frac{(1-m^\pi)}{2}(\pi_2-\pi^T) \tag{B.22}$$

$$(1+(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2})x=-\frac{\phi_1-1}{\sigma}\frac{(1-m^\pi)}{2}(\pi_2-\pi^T) \tag{B.23}$$

$$\pi = \pi_1 = \beta\frac{(1+m^\pi)}{2}\pi^T + \beta\frac{(1-m^\pi)}{2}\pi_2 - \kappa\frac{\frac{\phi_1-1}{\sigma}\frac{(1-m^\pi)}{2}(\pi_2-\pi^T)}{(1+(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2})} \tag{B.24}$$

$$\pi - \beta\pi^T = -(\pi-\pi^T)(\beta\frac{(1-m^\pi)}{2} - \kappa\frac{\frac{\phi_1-1}{\sigma}\frac{(1-m^\pi)}{2}}{(1+(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2})}) \tag{B.25}$$

$$(\pi-\pi^T)(1+\beta\frac{(1-m^\pi)}{2} - \kappa\frac{\frac{\phi_1-1}{\sigma}\frac{(1-m^\pi)}{2}}{(1+(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2})}) = 0 \tag{B.26}$$

with

$$m^x = tanh(\frac{b}{2}x_1^2) = tanh(-\frac{b}{2}x_2^2) \tag{B.27}$$

$$m^\pi = tanh(\frac{b}{2}(\pi_1-\pi^T)^2) = tanh(-\frac{b}{2}(\pi_2-\pi^T)^2) \tag{B.28}$$

So a 2-cycle must satisfy

$$(1+\beta\frac{(1-m^\pi)}{2} - \kappa\frac{\frac{\phi_1-1}{\sigma}\frac{(1-m^\pi)}{2}}{(1+(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2})}) = 0 \tag{B.29}$$

$$2+\beta-\kappa\frac{\frac{\phi_1-1}{\sigma}}{(1+(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2})} = m^\pi(\beta-\kappa\frac{\frac{\phi_1-1}{\sigma}}{(1+(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2})}) \tag{B.30}$$

$$m^\pi = \frac{(2+\beta)(1+(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2}) - \kappa\frac{\phi_1-1}{\sigma}}{\beta(1+(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2}) - \kappa\frac{\phi_1-1}{\sigma}} \tag{B.31}$$

In a 2-cycle around the fundamental steady state fundamentalists make prediction errors, while naive agents do not. Naive agents use the observation from period $t-1$ to give a prediction about period $t+1$. Therefore, in a 2-cycle they make no prediction errors, while fundamentalists do make prediction errors. In a 2-cycle we must therefore have

$$-1 \leq m_t^\pi < 0 \tag{B.32}$$

Now, if

$$(1+(1-\frac{\phi_2}{\sigma})\frac{(1-m_t^x)}{2})>0, \tag{B.33}$$

the inequalities of (B.32) reduce to

$$1+(1+\beta)\frac{\sigma}{\kappa}(1+(1-\frac{\phi_2}{\sigma})\frac{(1-m_t^x)}{2})\leq\phi_1<1+(2+\beta)\frac{\sigma}{\kappa}(1+(1-\frac{\phi_2}{\sigma})\frac{(1-m_t^x)}{2}) \tag{B.34}$$

If

$$(1+(1-\frac{\phi_2}{\sigma})\frac{(1-m_t^x)}{2})=0, \tag{B.35}$$

(B.32) can never hold, and if

$$(1+(1-\frac{\phi_2}{\sigma})\frac{(1-m_t^x)}{2})<0. \tag{B.36}$$

(B.32) reduces to

$$1+(2+\beta)\frac{\sigma}{\kappa}(1+(1-\frac{\phi_2}{\sigma})\frac{(1-m_t^x)}{2})<\phi_1\leq1+(1+\beta)\frac{\sigma}{\kappa}(1+(1-\frac{\phi_2}{\sigma})\frac{(1-m_t^x)}{2}) \tag{B.37}$$

As ϕ_1 comes close to making the right hand side of (B.34) or the left hand side of (B.37) binding, the system comes close to the fundamental steady state. In the limit we therefore have $m^x=0$. The limiting value of these restrictions therefore reduces to the bifurcation value

$$\phi_1^{PD}=1+(2+\beta)\frac{\sigma}{\kappa}(1+(1-\frac{\phi_2}{\sigma})\frac{1}{2})=1+(2+\beta)\frac{(3\sigma-\phi_2)}{2\kappa} \tag{B.38}$$

Finally, we can conclude that the bifurcation is subcritical (with a 2-cycle below the bifurcation) value if (B.33) holds for $m^x=0$, which is the case if and only if

$$\phi_2<3\sigma, \tag{B.39}$$

The bifurcation is supercritical (with a 2-cycle below the bifurcation) if

$$\phi_2>3\sigma, \tag{B.40}$$

and if $\phi_2=3\sigma$ the bifurcation occurs at $\phi_1=1$, and no 2-cycle is created.

B.5 Proof Proposition 4

The dynamical system given by (16) and (17) is linear in expectation fractions (m_t^x and m_t^π). Furthermore, since the system with all fundamentalists is degenerate (a steady state is reached in every period), the Jacobian and eigenvalues for any given set of expectation fractions is scaled by the fraction of naive agents. It follows that if the linear system given by (16) and (17) is stable for all naive fractions, it is stable for any set of expectations fractions, which implies global stability of the fundamental steady state in our non-linear dynamical system. When $m_t^x = m_t^\pi = -1$ the system reduces to

$$x_t = (1 - \frac{\phi_2}{\sigma})x_{t-1} - \frac{\phi_1 - 1}{\sigma}(\pi_{t-1} - \pi^T) \tag{B.41}$$

$$\pi_t = \beta\pi_{t-1} + \kappa x_t \tag{B.42}$$

The eigenvalues of this system are two times the eigenvalues given by (22) and (23). Now, replacing $\frac{1}{4}$ with $\frac{1}{2}$ in (B.6) and (B.17) and performing the same calculations as done in Appendix B.3 and B.4 gives the conditions given in the proposition.

B.6 Proof Proposition 5

Since $\phi_1^{opt} > 1$, the first condition from Proposition 4 is satisfied. The other condition for global stability reduces for $\phi_2 = \phi_2^{opt} = \sigma$ to

$$\phi_1^{opt} = 1 + \frac{\sigma\kappa}{\mu + \kappa^2} < 1 + (1 + \beta)\frac{\sigma}{\kappa}, \tag{B.43}$$

which is always satisfied for nonnegative μ.

C Zero lower bound on the nominal interest rate

C.1 Proof Proposition 6

From (28) and (29) it follows that in steady state, in the limit of $\beta \to 1$, we must have

$$\bar{x} = \frac{(1+m^{\pi})\pi^T + (1-m^{\pi})\bar{\pi}}{\sigma(1+m^x)}, \tag{C.1}$$

and

$$(\frac{1+m^{\pi}}{2} - \frac{\kappa}{\sigma}\frac{1-m^{\pi}}{1+m^x})\bar{\pi} = (\frac{1+m^{\pi}}{2} + \frac{\kappa}{\sigma}\frac{1+m_t^{\pi}}{1+m^x})\pi^T \tag{C.2}$$

In a steady state in the ZLB region with infinite intensity of choice differences in fractions can either be 0 or -1. Differences in fractions of 1 are not possible since in a steady state naive agents never make prediction errors. It follows that for infinite intensity of choice and for $\pi^T \neq 0$ the only possible solution of the above steady state equations is $\pi = x = m^x = 0$ and $m^{\pi} = -1$. This steady state only exists if it lies inside the ZLB region, which is the case if and only if

$$\begin{aligned} 0 &\leq (1 - \frac{1}{\phi_1})\pi^T \\ \phi_1 &\geq 1 \end{aligned} \tag{C.3}$$

Next, we turn to the stability of this steady state. The Jacobian evaluated at the steady state is

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{\sigma} & 0 & 0 & 0 & \frac{\pi^T}{2\sigma} \\ \frac{\kappa}{2} & (1+\frac{\kappa}{\sigma}) & 0 & 0 & 0 & (1+\frac{\kappa}{\sigma})\frac{\pi^T}{2} \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ c_{11}s_A & c_{12}s_A & 0 & 0 & 0 & e_{12}s_A \\ c_{21}s_B & c_{22}s_B & 0 & 0 & 0 & e_{22}s_B \end{pmatrix}$$

$$s_A = \frac{b}{2} sech(\frac{b}{2}A) \quad s_B = \frac{b}{2} sech(\frac{b}{2}B)$$

Where, $c_{11}, c_{21}, e_{12}, e_{22}, A$ and B are finite nonzero terms. If we let the intensity of choice, b, go to infinity, s_A and s_B go to zero. This means that the system has 4 eigenvalues equal to zero at the steady state. The other two follow from

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{\sigma} \\ \frac{\kappa}{2} & 1 + \frac{\kappa}{\sigma} \end{pmatrix}$$

and are given by

$$\lambda_1 = \frac{1}{2}[\frac{3}{2} + \frac{\kappa}{\sigma} - \sqrt{(\frac{3}{2} + \frac{\kappa}{\sigma})^2 - 2}] \tag{C.4}$$

$$\lambda_2 = \frac{1}{2}[\frac{3}{2} + \frac{\kappa}{\sigma} + \sqrt{(\frac{3}{2} + \frac{\kappa}{\sigma})^2 - 2}] \tag{C.5}$$

This means that the liquidity trap steady state is an unstable saddle point for all positive values of κ and σ ($\lambda_2 > 1$ and $|\lambda_1| < 1$ always hold).

C.2 Proof Proposition 8

A deflationary spiral (or divergence) is defined as a situations with ever decreasing inflation and output gap. If from any set of initial conditions in period t a deflationary spiral occurs, we must therefore at some future period $s \geq t$ have $x_{s+1} < x_s < x_{s-1} < x_{s-2} < 0$ and $\pi_{s+1} < \pi_s < \pi_{s-1} < \pi_{s-2} < 0$. From this it follows that naive agents turned out to perform better in their predictions about period s and $s+1$ than fundamentalists, so that, for infinite intensity of choice, we get $m^{\pi}_{s+1} = -1$, $m^{x}_{s+1} = -1$, $m^{\pi}_{s+2} = -1$ and $m^{x}_{s+2} = -1$

C.3 Recovery and divergence

To give some intuition in when recovery and divergence might occur, two lemmas are presented below.

Lemma 1 (**Low, positive inflation**). *If, at any time t, inflation and output gap realizations have led output gap expectations formed in the next period to be nonnegative ($E_{t+1}x_{t+2} \geq 0$), and inflation expectations formed in the next period to be positive ($E_{t+1}\pi_{t+2} > 0$), then, for infinite intensity of choice, recovery occurs.*

Proof. For $E_{t+1}\pi_{t+2} \geq \pi^T$ and $E_{t+1}x_{t+2} \geq 0$ the ZLB is not binding in period $t+1$, and recovery is trivial. We therefore need to show that recovery occurs when $m^{\pi}_{t+1} = -1$, $0 < \pi_t < \pi^T$, and either $m^x_{t+1} = 1$, or $m^x_{t+1} = -1$ and $x_t \geq 0$. It follows from (28) and (29) that in this case $x_{t+1} > 0$ and $\pi_{t+1} > \pi_t > 0$. Therefore the conditions of the proposition hold again in period $t+2$ implying that inflation will keep increasing. This reasoning can be continued until at some point in time inflation is high enough to make sure that (27) does no longer hold and recovery has occurred. □

Lemma 2 (**Deflation**). *If, at any time t, inflation and output gap realizations have led output gap expectations formed in the next period to be non-positive ($E_{t+1}x_{t+2} \leq 0$), and inflation expectations formed in the next period to be negative ($E_{t+1}\pi_{t+2} < 0$), then, for infinite intensity of choice, a sufficient condition for divergence is that inflation expectations will be naive in period $t+2$ ($m^{\pi}_{t+2} = -1$).*

Proof. The conditions in the proposition require that $m^{\pi}_{t+1} = -1, \pi_t < 0$, and either $m^x_{t+1} = -1$ and $x_t \leq 0$, or $m^x_{t+1} = 1$. From (28) and (29) it follows that when the above is satisfied $x_{t+1} < 0$ and $\pi_{t+1} < \pi_t < 0$. Now if $m^{\pi}_{t+2} = -1$, the above implies that in period $t+2$ the same conditions for negative inflation expectations and non positive output gap expectations are satisfied. It then follows from (28) and (29) that $x_{t+2} < 0$ and $\pi_{t+2} < \pi_{t+1} < \pi_t < 0$. The latter implies that naive agents performed better in their inflation prediction about period $t+2$ than fundamentalists, so that $m^{\pi}_{t+3} = -1$. The same reasoning can be continued from which we can conclude that inflation keeps decreasing. It then follows from (28) that at some point output gap expectations will become naive and output gap keeps decreasing as well □

From Lemma 1 it follows that if the economy is in a liquidity trap because of low past inflation, it will recover as long as there was no deflation, and agents do not expect a negative output gap. From Lemma 2 we can conclude that if the economy is in a liquidity trap because of deflation and a loss of credibility of the central bank with respect to inflation, there is a large danger of entering a liquidity trap with a self-fulfilling deflationary spiral. The only way that such a trap might be avoided is with a positive output gap together with limited credibility of the central bank with respect to output gap, so that output gap expectations are positive.

If we combine Lemma 1 and Lemma 2 we can conclude that when next periods output gap expectations are fundamentalistic, and inflation expectations are naive for the next *two* periods, recovery occurs if and only if there is no deflation $\pi_t > 0$. This is stated in the bottom two entries of the third column of Table 2.

C.4 Proof Proposition 9

Consider the limit of $\beta \to 1$. At least for the first two periods dynamics are given by the all naive system of

$$x_{t+1} = x_t + \frac{\pi_t}{\sigma} \tag{C.6}$$

$$\pi_{t+1} = (1 + \frac{\kappa}{\sigma})\pi_t + \kappa x_t \tag{C.7}$$

Iterating one more period gives

$$x_{t+2} = (1 + \frac{\kappa}{\sigma})x_t + \frac{2 + \frac{\kappa}{\sigma}}{\sigma}\pi_t \tag{C.8}$$

$$\pi_{t+2} = (1 + 3\frac{\kappa}{\sigma} + \frac{\kappa^2}{\sigma^2})\pi_t + (2\kappa + \frac{\kappa^2}{\sigma})x_t \tag{C.9}$$

The stable eigenvector of the all naive system that goes through the $x = \pi = 0$ steady state is given by

$$x_t = -\frac{1 + \sqrt{1 + 4\frac{\sigma}{\kappa}}}{2\sigma}\pi_t \tag{C.10}$$

As long as expectations remain all naive, inflation above this level leads output and inflation to move to the positive interest rate region and lower inflation leads to divergence to minus infinity for both variables.

In case of $x_t > 0$ and $\pi_t < 0$, output gap expectations that become fundamentalistic in period $t+3$ could cause divergence where recovery would have taken place in the all naive system. This cannot happen when

$$\begin{aligned} x_{t+2} = (1+\frac{\kappa}{\sigma})x_t + \frac{2+\frac{\kappa}{\sigma}}{\sigma}\pi_t &> \frac{x_t}{2} \\ x_t &> -\frac{4\sigma+2\kappa}{2\sigma\kappa+\sigma^2}\pi_t \end{aligned} \tag{C.11}$$

For $x_t < 0$ and $\pi_t > 0$ it follows from Lemma 1 that recovery occurs when output gap expectations become fundamentalistic, even when recovery would not have occurred in the all naive system. From

$$x_{t+2} = (1+\frac{\kappa}{\sigma})x_t + \frac{2+\frac{\kappa}{\sigma}}{\sigma}\pi_t > \frac{x_t}{2} \tag{C.12}$$

it can be seen that this happens exactly when (C.11) is satisfied.

It follows that when inflation expectations remain naive, recovery occurs if both π_t is larger than the value given by (C.10), and (C.11) is satisfied. This is summarized in the following condition

$$x_t > -max(\frac{1+\sqrt{1+4\frac{\sigma}{\kappa}}}{2\sigma}, \frac{4\sigma+2\kappa}{2\sigma\kappa+\sigma^2})\pi_t \tag{C.13}$$

Finally we must consider the possibility that inflation expectations become fundamentalistic. If $\pi_t < \pi^T$ fundamentalistic expectations only increase both variables for all subsequent periods, implying that (C.13) still is sufficient for recovery. If $\pi_t > \pi^T$ inflation expectations are at least as large as π^T, making the following condition sufficient for recovery.

$$x_t > -max(\frac{1+\sqrt{1+4\frac{\sigma}{\kappa}}}{2\sigma}, \frac{4\sigma+2\kappa}{2\sigma\kappa+\sigma^2})\pi^T \tag{C.14}$$

Putting Condition (C.13) and (C.14) together for the relevant values of π_t gives the expression in the proposition.

Next, we turn to sufficient conditions for divergence. If $\pi_t > \pi^T$ expectations becoming fundamentalistic lower inflation expectations, so this cannot prevent divergence when this would have occurred in the all naive system. It then follows from the above that in this case a sufficient condition for divergence to take place is

$$x_t < -max(\frac{1+\sqrt{1+4\frac{\sigma}{\kappa}}}{2\sigma}, \frac{4\sigma+2\kappa}{2\sigma\kappa+\sigma^2})\pi_t \tag{C.15}$$

If $\pi_t \leq \pi^T$ a sufficient condition for divergence is that both (C.15) holds and inflation expectations remain naive. The latter happens if and only if

$$\begin{aligned} \pi^T - \pi_{t+2} &> \pi_{t+2} - \pi_t \\ \pi_t &< \frac{\sigma^2\pi^T - (4\kappa\sigma^2 + 2\kappa^2\sigma)x_t}{\sigma^2 + 6\kappa\sigma + 2\kappa^2} \end{aligned} \tag{C.16}$$

C.5 Proof Proposition 10

For $b = 0$ the system (again considering the limit of $\beta \to 1$) then reduces to

$$x_{t+1} = \frac{x_t}{2} + \frac{\pi_t}{2\sigma} + \frac{\pi^T}{2\sigma} \tag{C.17}$$

$$\pi_{t+1} = (1+\frac{\kappa}{\sigma})(\frac{\pi_t}{2} + \frac{\pi^T}{2}) + \frac{\kappa}{2}x_t \tag{C.18}$$

The unique steady state of this system is

$$\pi = \frac{\sigma+2\kappa}{\sigma-2\kappa}\pi^T \tag{C.19}$$

$$x = \frac{2}{\sigma-2\kappa}\pi^T \tag{C.20}$$

The Jacobian of this linear system does not depend on the values of x and π and is given by

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2\sigma} \\ \frac{\kappa}{2} & \frac{1}{2}+\frac{\kappa}{2\sigma} \end{pmatrix} \tag{C.21}$$

Since the Jacobian is $\frac{1}{2}$ times the Jacobian of the system with all naive agents (given by (C.7) and (C.7)), it has the same eigenvectors, and the eigenvalues are given by

$$\lambda_1 = \frac{1}{2}[1 + \frac{\kappa}{2\sigma} - \sqrt{(1 + \frac{\kappa}{2\sigma})^2 - 1}] \tag{C.22}$$

$$\lambda_2 = \frac{1}{2}[1 + \frac{\kappa}{2\sigma} + \sqrt{(1 + \frac{\kappa}{2\sigma})^2 - 1}] \tag{C.23}$$

Both eigenvalues lie in the unit circle if and only if $\kappa < \frac{\sigma}{2}$, otherwise the steady state is a saddle point. In that case the slope of the stable eigenvector is given by (C.10).

References

Anufriev, Mikhail, Tiziana Assenza, Cars Hommes and Domenico Massaro (2013), 'Interest rate rules and macroeconomic stability under heterogeneous expectations'.

Arifovic, Jasmina and Luba Petersen (2015), Escaping expectations-driven liquidity traps: Experimental evidence. Manuscript, Simon Fraser University.

Assenza, Tiziana, Peter Heemeijer, Cars Hommes and Domenico Massaro (2014), Managing self-organization of expectations through monetary policy: a macro experiment, Technical report, CeNDEF Working Paper, University of Amsterdam.

Benhabib, Jess, George W Evans and Seppo Honkapohja (2014), 'Liquidity traps and expectation dynamics: Fiscal stimulus or fiscal austerity?', *Journal of Economic Dynamics and Control* **45**, 220–238.

Benhabib, Jess, Stephanie Schmitt-Grohé and Martin Uribe (2001*a*), 'Monetary policy and multiple equilibria', *American Economic Review* **91**(1), 167–186.

Benhabib, Jess, Stephanie Schmitt-Grohé and Martin Uribe (2001*b*), 'The perils of taylor rules', *Journal of Economic Theory* **96**(1), 40–69.

Berardi, Michele and John Duffy (2007), 'The value of central bank transparency when agents are learning', *European Journal of Political Economy* **23**(1), 9–29.

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 A and Bruce McGough (2010), 'Dynamic predictor selection in a new keynesian model with heterogeneous expectations', *Journal of Economic Dynamics and Control* **34**(8), 1492–1508.

Branch, William A and George W Evans (2011), 'Monetary policy and heterogeneous expectations', *Economic Theory* **47**(2-3), 365–393.

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

Brock, William A, Cars H Hommes and Florian Wagener (2005), 'Evolutionary dynamics in markets with many trader types', *Journal of Mathematical Economics* **41**(1), 7–42.

Bullard, James and Kaushik Mitra (2002), 'Learning about monetary policy rules', *Journal of monetary economics* **49**(6), 1105–1129.

Calvo, Guillermo A (1983), 'Staggered prices in a utility-maximizing framework', *Journal of monetary Economics* **12**(3), 383–398.

Clarida, Richard, Jordi Galí and Mark Gertler (1999), 'The science of monetary policy: a new keynesian perspective', *Journal of Economic Literature* **37**(4), 1661–1707.

Clarida, Richard, Jordi Galí and Mark Gertler (2000), 'Monetary policy rules and macroeconomic stability: evidence and some theory', *Quarterly Journal of Economics* **115**, 147–180.

Cornea, Adriana, Cars Hommes and Domenico Massaro (2013), Behavioral heterogeneity in us inflation dynamics, Technical report, Tinbergen Institute Discussion Paper.

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

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 (2003), 'Expectations and the stability problem for optimal monetary policies', *The Review of Economic Studies* **70**(4), 807–824.

Evans, George W and Seppo Honkapohja (2005), 'Policy interaction, expectations and the liquidity trap', *Review of Economic Dynamics* **8**(2), 303–323.

Evans, George W and Seppo Honkapohja (2009), 'Learning and macroeconomics', *Annu. Rev. Econ.* **1**(1), 421–449.

Galí, Jordi (2002), New perspectives on monetary policy, inflation, and the business cycle, Technical report, National Bureau of Economic Research.

Gomes, Orlando (2006), 'Optimal monetary policy under heterogeneous expectations', *The ICFAI Journal of Monetary Economics* **4**(2), 32–51.

Hommes, Cars, Domenico Massaro and Isabelle Salle (2015), 'Monetary and fiscal policy design at the zero lower bound–evidence from the lab'.

Hommes, Cars H (2013), *Behavioral rationality and heterogeneous expectations in complex economic systems*, Cambridge University Press.

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

Kurz, Mordecai, Giulia Piccillo and Howei Wu (2013), 'Modeling diverse expectations in an aggregated new keynesian model', *Journal of Economic Dynamics and Control* **37**(8), 1403–1433.

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

Manski, Charles F, Daniel McFadden et al. (1981), *Structural analysis of discrete data with econometric applications*, Mit Press Cambridge, MA.

Massaro, Domenico (2013), 'Heterogeneous expectations in monetary dsge models', *Journal of Economic Dynamics and Control* **37**(3), 680–692.

McCallum, Bennett T and Edward Nelson (1999), Performance of operational policy rules in an estimated semiclassical structural model, *in* 'Monetary policy rules', University of Chicago Press.

McCallum, Bennett T and Edward Nelson (2004), 'Timeless perspectives vs. discretionary monetary policy in forward-looking models', *Federal Reserve Bank of St. Louis Review* **86**(2), 43–56.

Mertens, Karel RSM and Morten O Ravn (2014), 'Fiscal policy in an expectations-driven liquidity trap', *Review of Economic Studies* **81**(4), 1637–1667.

Milani, Fabio (2007), 'Expectations, learning and macroeconomic persistence', *Journal of monetary Economics* **54**(7), 2065–2082.

Orphanides, Athanasios (2004), 'Monetary policy rules, macroeconomic stability, and inflation: A view from the trenches', *Journal of Money, Credit and Banking* pp. 151–175.

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

Rotemberg, Julio J and Michael Woodford (1999), Interest rate rules in an estimated sticky price model, *in* 'Monetary policy rules', University of Chicago Press, pp. 57–126.

Salle, Isabelle, Murat Yıldızoğlu and Marc-Alexandre Sénégas (2013), 'Inflation targeting in a learning economy: An abm perspective', *Economic Modelling* **34**, 114–128.

Svensson, Lars E.O. (2003), 'What is wrong with taylor rules? using judgment in monetary policy through targeting rules', *Journal of Economic Literature* **41**(2), 427–477.

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

Taylor, John B (1999), A historical analysis of monetary policy rules, *in* 'Monetary policy rules', University of Chicago Press, pp. 319–348.

Walsh, Carl E (2003), 'Speed limit policies: the output gap and optimal monetary policy', *American Economic Review* **93**, 265–278.

Woodford, Michael (1999), 'Optimal monetary policy inertia', *The Manchester School* **67**(s1), 1–35.

Woodford, Michael (2002), 'Inflation stabilization and welfare', *Contributions in Macroeconomics* **2**(1).

Woodford, Michael (2003), *Interest and prices: Foundations of a theory of monetary policy*, Princeton University Press.

Woodford, Michael (2004), 'Inflation targeting and optimal monetary policy', *Federal Reserve Bank of St. Louis Economic Review* **86**(4), 15–41.