Managing Unanchored, Heterogeneous Expectations and Liquidity Traps*

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

[a]CeNDEF, University of Amsterdam

[b]Tinbergen Institute, Amsterdam

August 14, 2017

Abstract

We study the possibility of (almost) self-fulfilling waves of pessimism and self-reinforcing liquidity traps in a New Keynesian model with heterogeneous expectations. We explicitly focus on the "anchoring" of expectations that is modeled as the range of deviations from the central bank targets (and from the rational expectation equilibrium) that agents are willing to consider. We find that when the zero lower bound on the nominal interest rate is not binding, aggressive monetary policy can prevent waves of pessimism and exclude near unit root dynamics, even when expectations are unanchored. However, as shocks bring the economy to a situation with a binding zero lower bound, there is a danger of a long lasting self-reinforcing liquidity trap that arises because of the existence of multiple steady states. It turns out that in a model where the anchoring of expectations evolves endogenously, the anchoring of expectations at the time the bad shocks hit is crucial in determining whether the economy can recover from the liquidity trap. Furthermore, a higher inflation target reduces the probability that self-reinforcing liquidity traps arise.

JEL Classification: E52, E32, C62

Keywords: Interest Rate Rules, Liquidity Traps, Heterogeneous Expectations, Bounded Rationality, Multiple Steady States

---

*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

The aftermath of the recent financial crises has highlighted the importance of the zero lower bound on the nominal interest rate and the possibility of self-reinforcing liquidity traps. Interest rates all over the world have been at low levels for a prolonged period of time, while recovery of output and inflation has been slow. While many factors may play a role in explaining this, low confidence is likely to be an important one. As consumers expect slow economic growth and a high risk of unemployment in the future, they are reluctant to increase their current consumption. This keeps aggregate demand low, and hence prevents a recovery in output. This problem is amplified by the fact that a combination of low inflation and near zero nominal interest rates leads to a high real interest rate, which further reduces consumption. As a consequence, output stays even lower, further reducing confidence. The economy then is stuck in a liquidity trap with a binding zero lower bound and low output and inflation.

An important reason for increased transparency of many central banks and the adoption of inflation targeting, has been to *anchor* private sector expectations to a positive target value. While this policy has initially been very successful in stabilizing inflation, it seems that since the financial crisis private sector expectations have no longer been anchored to the target. Instead, many market participants consider the possibility that inflation will stay well below target for a prolonged period of time. Additionally, output expectations are no longer anchored to their long run growth path, and economists worry that economies might not rerun to this growth path any time soon. These unanchored expectations are crucial for explaining how a self-reinforcing liquidity trap as described above could arise.

While the widely used rational expectations paradigm may be useful for many applications, we believe that it is lacking in its ability to model dropping confidence and changes in the anchoring of expectations in a satisfactory way. Some successful attempts have been made at explaining expectations driven liquidity traps by deviating, to some extent, from the traditional paradigm. Mertens and Ravn (2014) build a model where sunspot shocks can drive the economy temporarily towards a low inflation low output equilibrium with a binding zero lower bound. The existence of such a zero lower bound equilibrium was first highlighted by Benhabib et al. (2001*a*,*b*). Evans et al. (2008), Eusepi (2007, 2010), and Benhabib et al. (2014) take a different approach and study dynamics in a model where agents use adaptive learning to form expectations. In this framework they find that deflationary spirals with ever decreasing inflation and output can arise. In all these models, expectations are however homogeneous and, therefore, either in line with the rational expectations equilibrium consistent with the central banks targets, or deviating from this equilibrium.

We believe that more insight in the possibility of self-fulling liquidity traps can be obtained by allowing for *heterogeneity* in expectations in a framework where we can explicitly model the extent to which these heterogeneous expectations are *anchored* to the targets of the central bank and the corresponding desired rational expectations equilibrium. Besides giving a more intuitive notion to the anchoring of expectations, heterogeneous expectations are also arguably more realistic than the assumption of a representative agent. Both surveys of consumers and professional forecasters (e.g. Mankiw et al. (2003); Carroll (2003); Andrade et al. (2014)) and laboratory experiments with human subjects (e.g. Pfajfar and Zakelj (2011); Assenza et al. (2014)) show that in practice there exists considerable heterogeneity in the forecasts of macroeconomic variables.

In our model, agents consider a distribution of expectation values around the targets of the central bank and the desired rational expectations equilibrium. When this distribution of expectation values has a small variance, we say that expectations are strongly anchored to the targets. When the variance is large and agents also consider expectation values that are far away from the targets, expectations are unanchored. In each period, each agent picks an expectation from the distribution of possible expectation values. They do this randomly, where the probability of choosing each expectation value depends on past realizations. That is, when agents that had pessimistic expectations turn out to be right (e.g. due to bad shocks to the economy), in subsequent periods, other agents are more likely to choose pessimistic expectations as well. However, there will always be heterogeneity present, with some agents expecting values close to the targets and other agents having more pessimistic or optimistic expectations.

We find that *self-reinforcing waves of pessimism* can arise, where agents coordinate on expectations below target. These low expectations then result in low inflation and output gap realizations, which in turn reinforce the pessimistic expectations. In normal times, such waves of pessimism can be limited, or even prevented, with aggressive enough monetary policy that provides strong mean reversion to the target steady state. Inflation and output gap will then remain relatively close to their targets, even when expectations are low. However, as the economy is driven to the zero lower bound on the nominal interest rate, monetary policy is no longer able to provide strong mean reversion, and self-fulfilling waves of pessimism can arise more easily. These can then only be prevented if expectations are relatively strongly anchored to the targets, and the inflation target is high enough.

We first illustrate the above basic mechanism in a styled, analytically tractable model with only 3 types of agents: optimist, pessimist and fundamentalists. The latter expect inflation and output gap to be exactly at their targets, while optimist (pessimists) have somewhat higher (lower) expectations. In this model specification, the anchoring of expectations is determined by how much above or below target optimists and pessimists expect variables to be. We show that if pessimists have low enough expectations relative to the inflation target, agents can coordinate on a non-fundamental *pessimistic liquidity trap steady* state with a binding zero lower bound.

We then turn to our benchmark model where there are many different expectation values, each spaced by 0.5%. This discreteness in expectation formation reflects that, in practice, people do not report expectations with infinitely many decimals, but instead prefer round numbers (see e.g. Assenza et al. (2014)).[1] In this model, the anchoring of expectations, which is given by the number of expectation types (and hence the highest and lowest expectation values that agents consider), is allowed to vary *endogenously* with past absolute deviations from target in the economy. We show that, in this model, multiple optimistic and pessimistic steady states can co-exist even in the absence of a zero lower bound on the nominal interest rate. This is however only the case if monetary policy is very weak so that there is so little mean reversion that expectations become almost self-fulfilling.

As in the 3-type model, introducing the zero lower bound makes it much more plausible that a self-reinforcing wave of pessimism can bring the economy to a *pessimistic liquidity trap steady state*. Such a wave of pessimism could e.g. be triggered by a couple of bad shocks to the fundamentals of the economy. We show that the anchoring of expectations at the time that the economy is hit by bad shocks is crucial in determining whether the economy can recover, or instead ends up in a liquidity trap that can last for prolonged

[1]This phenomenon is labeled "digit preference". See Curtin (2010).

periods of time. In both cases the bad shocks initiate a wave of pessimism where the majority of agents coordinate on the lowest expectations that they are willing to consider, given the current anchoring of expectations. If the anchoring is strong enough, then these expectations will not be too low, and small positive shocks to economic fundamentals can bring the economy back on a path to the target steady state. If expectations are more unanchored and the wave of pessimism is more severe, larger shocks would be needed to achieve this. If such larger shocks do not arise, the economy will remain in the liquidity trap, which slowly becomes worse as expectations become less anchored over time. We conclude that the central bank must make sure that it anchors expectations strongly to its targets during times where the zero lower bound is not binding. Furthermore, a high enough inflation target could prevent such liquidity traps.

Finally, we use a model with a continuum of heterogeneous expectation rules, that gives very similar dynamics to our benchmark model, to provide analytical results. We do this by applying the large type limit (LTL) concept of Brock et al. (2005) to the New Keynesian framework. The largest difference between this model and our benchmark model arises because in the LTL specification there is no lower bound on expectation values that agents consider, so that coordination on a pessimistic steady state cannot arise. Instead, a self reinforcing wave of pessimism takes the form of a *deflationary spiral* as in Evans et al. (2008), Benhabib et al. (2014) and Eusepi (2010) and other with adaptive learning. With our LTL model we add to that literature by presenting analytical expressions for the initial conditions that lead to a deflationary spiral or to recovery (what they call a "corridor of stability"). Moreover, our benchmark model adds to the above literature by creating the possibility that downward spiraling expectations do not continue indefinitely but instead lead to converge to a locally stable liquidity trap steady state. This is also the case for Evans et al. (2016). Where however in that paper exogenous lower bounds on inflation and consumption are imposed in order to create a locally stable steady state under the zero lower bound, such a liquidity trap steady state arises in our model because of anchored expectations. This steady state can furthermore shift endogenously as the anchoring of expectations evolves over time. Finally, where the adaptive learning literature assumes homogeneous expectations, we extend the analysis by allowing for heterogeneity in expectations.

Our model is further related to the heterogeneous expectation models of Branch and

McGough (2010), Massaro (2013), Gasteiger (2014) and Di Bartolomeo et al. (2015). These papers however assume that agents can choose between two expectation formation rules, rather than a distribution of heterogeneous expectations. More closely related are the works of Anufriev et al. (2013), De Grauwe (2011), Agliari et al. (2014), Agliari et al. (2015) and Pecora and Spelta (2017). However, these papers discuss either a model with a relatively small number of different expectation values (like the 3-type model), or the large type limit specification with a continuum of expectation values. None of these papers consider the case where the range of expectation values varies endogenously, and where the number of expectation values can be large but finite.

Moreover, where the focus in the above literature is on the desired, and in some cases optimal, specification of monetary policy when the zero lower bound is not binding, our main goal is to investigate how self-reinforcing waves of pessimism can arise when conventional monetary policy can no longer be used due to the restrictions of the zero lower bound on the nominal interest rate. De Grauwe and Ji (2016) and Hommes and Lustenhouwer (2015) consider the zero lower bound in an heterogeneous expectations model. Our contribution differs form these papers in that De Grauwe and Ji (2016) rely on simulations only and do not derive analytical results, and Hommes and Lustenhouwer (2015) focuses on the credibility of the central bank rather than the anchoring of expectations, in a model with only two types. Furthermore, in neither of these papers locally stable liquidity trap steady states can arise.

The remainder of this paper is organized as follows. In Section 2 the economic model, and general expectation formation mechanism are presented. In Section 3 we illustrate with the stylized 3-type example how (almost) self-fulfilling waves of pessimism could arise, both without and with the zero lower bound on the nominal interest rate. Section 4 considers our benchmark model with an endogenously varying number of types, as well as the large type limit with a continuum of expectation values. Section 5 concludes.

2 Model specification

2.1 Economic environment

We use a log linearized New Keynesian model in line with Woodford (2003). Microfoundations for this model when expectations are heterogeneous can be found in Hommes and Lustenhouwer (2015). The derivations in that paper largely follow Kurz et al. (2013).[2] Microfoundations of the same New Keynesian model with heterogeneous expectations under alternative assumptions are derived by Branch and McGough (2009).

The model is given by a New Keynesian Phillips curve describing inflation, an IS curve describing output gap, and a rule for the nominal interest rate. Log linearized output gap (x_t) and inflation (π_t) are given by

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

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

where κ, σ and β are model parameters, and $\bar{r} = \frac{1}{\beta} - 1$ is the steady state real interest rate. e_t and u_t are shocks to the economy, which we assume to be white noise.

Finally, i_t is the nominal interest rate.[3] We assume the nominal interest rate is set according to a forward-looking Taylor type interest rate rule.[4]

$$i_t = \bar{r} + \pi^T + \phi_1(E_t \pi_{t+1} - \pi^T) + \phi_2(E_t x_{t+1} - x^T), \tag{3}$$

where π^T is the central banks inflation target, and $x^T = \frac{1-\beta}{\kappa}\pi^T$ is the output gap target consistent with π^T. Forward-looking rules generally perform well empirically (e.g. Clarida et al., 1998). Furthermore, Evans and Honkapohja (2003) show that a forward-looking rule of the form (3), can be used to minimize an (add hoc) loss function containing discounted squared deviations of inflation and output gap from target.[5] Branch and

[2] In the derivations in Hommes and Lustenhouwer (2015), the aggregation of individual decisions in terms of aggregates only (which is not the case in Kurz et al. (2013)), hinges on a specific property of the way we model heterogeneous expectations. With the Heuristic switching model of Brock and Hommes (1997), in any period there is heterogeneity in expectations, with different fractions of agents using different heuristics. However, these fractions are updated in each period according to a probability distribution that depends on the relative past performance of each expectation formation heuristic. With this updating process each agent has in each period the same probability of following a particular heuristic as all other agents, independent of the heuristic the particular agent followed in the previous period. By assuming that agents are aware of this, their expectations about their own future consumption coincide with their expectation about the future consumption of any other agents, and therefore with their expectations about aggregate consumption, which is crucial for aggregation. See Hommes and Lustenhouwer (2015) for details.

[3] Note that different from Hommes and Lustenhouwer (2015), the nominal interest rate is here not defined in terms of deviation from steady state (which would give $\hat{\imath}_t = i_t - \bar{r}$). This makes the zero lower bound analysis in the current paper more intuitive.

[4] All results in this paper carry over qualitatively to the case of a contemporaneous Taylor rule or a Taylor rule with lagged data. These results can be found in Lustenhouwer (2017).

[5] We assume here that the central bank can not observe and respond to contemporaneous shocks. Otherwise these would show up as well in a rule derived from such a loss function.

Evans (2011) and Hommes and Lustenhouwer (2015) show that this "optimal" rule also works well in models with heterogeneous expectations.

Abstracting from shocks and plugging (3) into (1), gives the following model

$$x_t = (1 - \frac{\phi_2}{\sigma})E_t x_{t+1} - \frac{\phi_1 - 1}{\sigma}(E_t \pi_{t+1} - \pi^T) + \frac{\phi_2}{\sigma} x^T, \tag{4}$$

$$\pi_t = \beta E_t \pi_{t+1} + \kappa x_t. \tag{5}$$

2.2 Expectation formation

We deviate from the rational expectations hypothesis, and do not assume all agents always exactly expect the same outcome of future variables. Instead, we assume that some **heterogeneity** is present, with some agents expecting values that are a bit higher and some agents expecting values that are a bit lower. This heterogeneity could be caused by agents making small mistakes. Alternatively, the heterogeneity can arise because some agents think they have reasons to be more optimistic or pessimistic in their predictions than is warranted by the publicly available information. More specifically, expectations are *distributed around the targets of the central bank*, π^T an x^T. This can be interpreted as some agents trusting the central bank and expecting inflation and output gap to be equal to their targets, while other agents expect that the central bank will not be able to exactly achieve its goals, but that inflation and output gap will be somewhat higher or lower. An alternative interpretation is that expectations are heterogeneously distributed around the minimum state variable (MSV) solution of the model. Since we assume no autocorrelation in shocks, the MSV solution coincides with the targets of the central bank in any period.

We further assume that when agents that were more optimistic or pessimistic in their prediction turn out to be right, then other agents will *learn* from this and adjust their expectations in the direction of the better performing agents. Agents may, for example, think that the correct agents had additional information available to them, or just had more skills in analyzing the economic environment.

We implement this way of expectation formation with a heuristic switching model as in Brock and Hommes (1997), where agents switch between simple prediction rules, or heuristics. The heuristics in our model consists of deviations from the fundamental values of the economy. The fraction of agents using the heuristic with deviation, or bias, $b_{z,h}$ in period t is updated according to the discrete choice model with multinomial logit probabilities (see Manski et al. (1981)) given by

$$n_t^{z,h} = \frac{e^{\omega U_{t-1}^{z,h}}}{\sum_{h=1}^{H^z} e^{\omega U_{t-1}^{z,h}}}, \qquad z = \pi, x. \tag{6}$$

Here, H^π and H^x are the total number of prediction values used for respectively inflation and output gap. $U_t^{z,h}$ is the fitness measure of predictor h in period t, which we will depend on past prediction errors, and ω is the intensity of choice. The higher the intensity of choice, the more sensitive agents become with respect to relative performance of prediction values, and the more forecasts of agents will be coordinated.

Aggregate expectations are now given by a weighted average of the predictions of all types. This gives for respectively inflation and output gap

$$E_t \pi_{t+1} = \pi^T + \sum_{h=1}^{H^\pi} b_{\pi,h} n_t^{\pi,h}, \tag{7}$$

and

$$E_t x_{t+1} = x^T + \sum_{h=1}^{H^x} b_{x,h} n_t^{x,h}, \tag{8}$$

3 An example with 3 types

Before we turn to the richer model with a large number of prediction values in Section 4, we will first illustrate some key properties of our model with a simple example where $H^x = H^\pi = 3$. There then are three types of agents for both output and inflation, which we can label **fundamentalists**, **optimist** and **pessimists**. Fundamentalists have no bias and believe in the targets of the central bank (or alternatively in the MSV solution of the model). Their expectations are therefore given by $E_t^{fun} x_{t+1} = x^T$ and $E_t^{fun} \pi_{t+1} = \pi^T$. Then there are optimists and pessimists who have a bias of b and $-b$ respectively. Their expectations are given by $E_t^{opt} x_{t+1} = x^T + b$, $E_t^{opt} \pi_{t+1} = \pi^T + b$, $E_t^{pes} x_{t+1} = x^T - b$, and $E_t^{pes} \pi_{t+1} = \pi^T - b$. The magnitude of the bias b, determines the range of values around the targets (i.e. MSV solution) that agents are willing to consider as expectations. We take this as a measure of the *"anchoring of expectations"*. If b is small, then agents will always expect values close to the targets of the central bank. We say that expectations are **strongly anchored** in this case. If on the other hand b is large, then agents could potentially switch to expectations that lie far from the target by becoming optimistic or pessimistic. In this case, expectations are **unanchored**. In this section, we will assume that b is fixed, so that the anchoring of expectations is exogenously given. In Section 4 we will allow for endogenous evolution of the anchoring of expectations. Note that we allow an agent to be optimistic about one variable, while being pessimistic or fundamentalistic about the other, so that the fractions of agents that are optimistic and pessimistic (denoted respectively $n_t^{z,opt}$ and $n_t^{z,pes}$) may differ between the two variables $z = x, \pi$. The fraction of fundamentalists of variable z equals $n_t^{z,fun} = 1 - n_t^{z,opt} - n_t^{z,pes}$.

In this 3-type model, we will first show how our heterogeneous expectations framework with nonlinear updating can imply *multiple steady states*, even in an otherwise linear model (Section 3.1). Here we also illustrate how monetary policy can prevent the existence of multiple steady states, so that coordination on optimism or pessimism is no longer possible in the long run. In Section 3.2 we introduce the zero lower bound on the nominal interest rate and show that this drastically alters policy implications. Now the existence of a liquidity trap steady state with coordination on pessimism for both inflation and output can no longer be prevented by choosing adequate coefficients in the Taylor rule. Instead, such a "bad" equilibrium can only be ruled out with a sufficiently high inflation target or with sufficiently anchored expectations.

3.1 Steady states in the 3-type model without zero lower bound

We first ignore the zero lower bound on the nominal interest rate and consider the model given by (4) and (5). Plugging in the biases of fundamentalists, optimist and pessimist (respectively $b_{x,h} = -b$, $b_{x,h} = 0$, and $b_{x,h} = b$ for output, and $b_{\pi,h} = -b$, $b_{\pi,h} = 0$, and $b_{\pi,h} = b$ for inflation) in (7) and (8) results in

$$E_t x_{t+1} = x^T + b(n_t^{x,opt} - n_t^{x,pes}), \tag{9}$$

$$E_t \pi_{t+1} = \pi^T + b(n_t^{\pi,opt} - n_t^{\pi,pes}). \tag{10}$$

Fractions are given by (6), with $h = opt, pes, fun$. For analytical tractability, we will in this section assume that the fitness measures are given by the most recently observed squared prediction error. That is $U_{t-1}^{\pi,h} = -(\pi_{t-1} - E_{t-2}^h \pi_{t-1})^2$ for inflation, and $U_{t-1}^{x,h} = -(x_{t-1} - E_{t-2}^h x_{t-1})^2$ for output gap, with $h = opt, pes, fun$. In Section 4 we will allow the fitness measure to also depend on forecasting performance in earlier periods.

We now look at what steady states can exist in this non-linear model.

Proposition 1 (Existence fundamental steady state 3-type model). *A fundamental steady state where inflation and output gap are equal to the targets of the central bank, and where being a fundamentalist is the best performing heuristic, always exists, independent of monetary policy and the parametrization of the model.*

Proof. Plugging in the expectations of the three types in the above fitness measures, it follows that when $\pi_{t-1} = \pi^T$ and $x_{t-1} = x^T$, optimists and pessimists have identical fitness ($U_{t-1}^{x,opt} = U_{t-1}^{x,pes}$ and $U_{t-1}^{\pi,opt} = U_{t-1}^{\pi,pes}$). It then follows from (6) that in this case $n_t^{x,opt} = n_t^{x,pes}$ and $n_t^{\pi,opt} = n_t^{\pi,pes}$, so that (9) and (10) reduce to $E_t x_{t+1} = x^T$ and $E_t \pi_{t+1} = \pi^T$. Plugging these expectations in (4) and (5), realized inflation and output gap will, in the absence of shocks, be equal to $\pi_t = \pi^T$ and $x_t = x^T$. This will then be true in any period, so that the targets of the central bank comprise a steady state. ■

While Proposition 1 implies that, in the absence of shocks, the central bank could achieve its target in every period, there might also exist *additional steady states*, where inflation and output gap are not equal to the target. This possibility arises due to the non-linear updating process of the expectation fractions (6). In such non-fundamental steady states, fundamentalism would no longer be the best performing heuristic for both variables, but, instead, optimism or pessimism would dominate for inflation and/or output gap. Since agents can be either pessimist, fundamentalist, or optimist about both inflation *and* output gap, there are a total of nine combinations of dominating heuristics that could potentially comprise a steady state. That is, beside the fundamental steady state, eight different non-fundamental steady states could potentially exist.

As we are interested in the possibility of a liquidity trap, and this is most likely to arise under pessimistic expectations, we will focus on the steady state where pessimism dominates for both inflation and output gap. In Section 3.2 we will include the zero lower bound on the nominal interest rate and see whether this steady state can imply a liquidity trap.

Proposition 2 states under what conditions this pessimistic steady state exists for the limiting case where the intensity of choice goes to infinity ($\omega = \infty$). This implies that agents in every period immediately switch to the best performing heuristic, and coordinate perfectly on either optimism, pessimism or fundamentalism.[6] Proof of the proposition is given in Appendix A.1.

Proposition 2 (Existence pessimistic steady state 3-type model). *Without the zero lower bound on the nominal interest rate, a locally stable steady state were pessimism dominates for both inflation and output gap exists under $\omega = \infty$ if and only if $\phi_1 + \phi_2 < 1 + \frac{\sigma}{2}$.*

From Proposition 2 it follows that if the central bank does not respond to output gap (or at least not very strongly), then the pessimistic steady state can exist, in the limiting case of infinite intensity of choice[7], even if $\phi_1 > 1$. That is, even if the Taylor principle is satisfied, and the fundamental equilibrium would be locally determinate under rational expectations, the additional pessimistic steady state can exist. However, when the central bank responds more *aggressively* to inflation (and/or to output gap), the existence of the pessimistic steady state is ruled out.[8]

The intuition for the existence of the pessimistic steady state when monetary policy is not aggressively enough is the following. If agents are pessimists, then this induces inflation and output to become so low that agents have better predictive success by being pessimists, than by being fundamentalists. We call such expectations **"almost self-fulfilling"**. When, on the other hand monetary policy is aggressively enough, the central bank will sufficiently lower the interest rate when expectations become pessimistic, which will stimulate output and inflation, leading to realizations that are closer to the fundamental values. Pessimistic expectations then perform worse than fundamental expectations and the pessimistic steady state does not exist.

When pessimistic expectations are almost self-fulfilling, it could still be that pessimists

[6]The exception to perfect coordination arises in the knife edge case where fundamentalists have exactly the same fitness as optimists (pessimists). In this case half of the agents will become fundamentalists, while the other half become optimists (pessimists). These knife edge cases could also comprise steady states. However, we show in the online supplementary material that such steady states are always unstable.

[7]For *finite* intensity of choice there would not be perfect coordination, implying that during a wave of pessimism also some agents are fundamentalists and optimists. These agents would put upward pressure on inflation and output gap, driving realizations further away from the expectations of pessimists, making them less self-fulfilling. In this case the pessimistic steady state can still exist, but the condition for its existence would become more stringent.

[8]The other 7 non-fundamental steady states can also be ruled out with an aggressive enough response to inflation and/or output gap. Conditions for existence of these steady states can be found in Lustenhouwer (2017). In particular, the condition for existence of an optimistic steady state is exactly the same as the one given in Proposition 2, since (abstracting from the zero lower bound) the model is completely symmetric.

make significant forecast errors. What is key however, is that their forecast errors are *smaller* than those of optimists and fundamentalists and that pessimism outperforms all other available prediction values. When a larger number of expectation values is allowed (as in Section 4), forecast errors are required to be smaller for multiple steady states to exist. Below, we will also consider the case where pessimistic expectations lead to realizations that are not only low enough to make pessimistic expectations almost self-fulfilling, but actually strictly lower than the expectations ($z_t < E_{t-1}z_t$). We will refer to this case as **"self-reinforcing"** pessimistic expectations. Without the zero lower bound, a self-reinforcing wave of pessimism can only arise if monetary policy does not satisfy the Taylor principle.[9] However, when we take account of the zero lower bound on the nominal interest rate, self-reinforcing waves of pessimism are more likely to arise, as we will show next.

3.2 A liquidity trap steady state in the 3-type model

We now introduce the zero lower bound (ZLB) on the nominal interest rate in the above 3-type model, and investigate its consequences on the existence of the pessimistic steady state. In normal times the interest rate is still given by the forward-looking Taylor rule, (3). We then say that our model is in the "positive interest rate region". When, however, the Taylor rule implies that $i_t < 0$, then i_t is instead set equal to zero. In that case, we say that the system is in the "ZLB region", and inflation and output reduce to

$$x_t = E_t x_{t+1} + \frac{1}{\sigma} E_t \pi_{t+1} + \frac{\bar{r}}{\sigma}, \tag{11}$$

$$\pi_t = \beta E_t \pi_{t+1} + \kappa x_t. \tag{12}$$

Proposition 3 states the conditions for existence of the pessimistic steady state under the zero lower bound. Its proof is given in Appendix A.2

Proposition 3 (**Existence liquidity trap steady state 3-type model**). *When all agents are pessimists, the zero lower bound on the nominal interest rate is binding if and only if*

$$\pi^T + \bar{r} < b(\phi_1 + \phi_2). \tag{13}$$

[9]It would require $\phi_1 + \phi_2 < 1$.

In that case, the pessimistic steady state of proposition 2 becomes a liquidity trap steady state, and exists (assuming $\omega = \infty$) if and only if

$$\pi^T + \bar{r} < b(1 + \frac{\sigma}{2}). \tag{14}$$

From Proposition 3 it follows that a **liquidity trap steady state** with pessimistic expectations and a binding zero lower bound, can exist if expectations are *unanchored* (b large). However, its existence can then be prevented if the *inflation target* is sufficiently high. Note furthermore, that the liquidity trap steady state can exist when a pessimistic steady state would also have existed in the absence of a zero lower bound on the nominal interest rate ($\phi_1 + \phi_2 < 1 + \frac{\sigma}{2}$); but that it can also exist when monetary policy is aggressive enough to rule out the existence of a pessimistic steady state without the ZLB ($\phi_1 + \phi_2 > 1 + \frac{\sigma}{2}$). In the first case, (13) is the binding condition for the existence of the liquidity trap steady state, while in the second case it is (14).

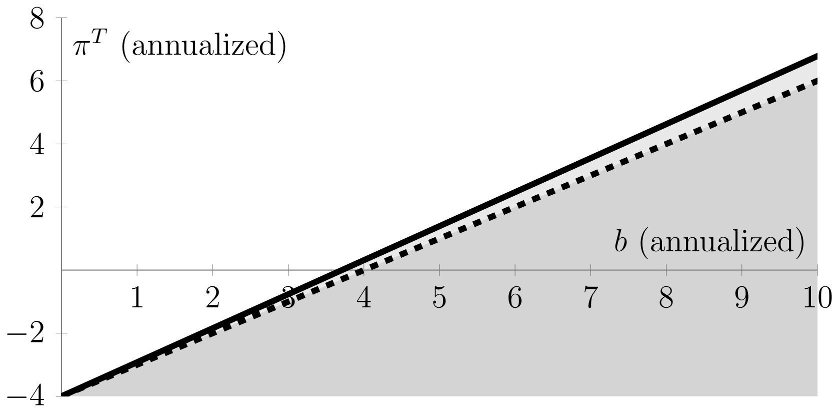

Figure 1: Existence of liquidity trap steady state of Proposition 3 when $\phi_1 + \phi_2 > 1 + \frac{\sigma}{2}$, $\sigma = 0.157$ and $\beta = 0.99$, depending on the anchoring of expectations (b) and the inflation target (π^T). In the shaded area below the solid line (including both the light and the dark part) the liquidity trap steady state exists, while in the non-shaded area above the black line it is ruled out. Below the dashed line (darker shaded area) pessimistic expectations are self-reinforcing.

In Figure 1, we plot the condition for existence of the liquidity trap steady state in case of $\phi_1 + \phi_2 > 1 + \frac{\sigma}{2}$ (i.e. we plot (14)) with a solid line (the meaning of the dashed line will be discussed below). Here, we calibrate the elasticity of inter temporal substitution as in Woodford (1999) to $\sigma = 0.157$, and set $\beta = 0.99$. For higher calibrations of σ the line in Figure 1 would be steeper, implying that a higher inflation target would be needed to rule out the liquidity trap steady state for a given level of the anchoring of expectations (b). In the figure, it can be seen that, as long as pessimists expect annualized inflation to be within 3.5% of its target ($b < 3.5$), the liquidity trap does not exist even if the central bank's inflation target is zero. However, as expectations become more unanchored, a considerably higher inflation target is needed to rule out the possibility of a pessimistic liquidity trap.

Finally, we turn to the possibility of *self-reinforcing* expectations in a liquidity trap. Remember that in order for the steady state to exist, expectations must at least be almost self-fulfilling in the sense that they result in realizations of inflation and output that are closer to the pessimistic expectations than to the fundamentalist ones. However, we now consider whether expectations can also lead to realizations of inflation and output gap that are as low or lower than what pessimists expected. Proposition 4, whose proof is given in Appendix A.3, states when this is the case.

Proposition 4 (**A self-reinforcing liquidity trap**). *The liquidity trap steady state of Proposition 3 is self-reinforcing if and only if*

$$\pi^T + \bar{r} < b \tag{15}$$

Condition (15) is plotted as a dashed line in Figure 1. It follows that for most of the combinations of anchoring of expectations (b) and the inflation target (π^T) where the pessimistic steady state exists, it is also self-reinforcing (the darker shaded area below the dashed line).

4 Many types

We now turn to our benchmark model with a large number of expectation values. While the 3-type model of Section 3.1 gives interesting insights (in part because of its analytical tractability), it is rather stylized in the sense that it does not allow for different degrees of optimism and pessimism. Instead, all pessimists always have the same expectations and all optimists always have the same expectations. The model in this section allows for different degrees of optimism and pessimism by introducing more possible expectation values. In addition, we will allow the number of expectation values that agents consider and their range (which we interpret as the anchoring of expectations in this model), to vary over time *endogenously.*

Even though it makes sense to assume that different degrees of optimism and pessimism are allowed, some **discreteness** in expectations is empirically relevant. In practice, people do not form expectations with infinitely many decimals. Instead, humans prefer round numbers when reporting their expectations (a phenomenon labeled *digit preference*).[10] We choose to model this by assuming that agents form expectations with a precision of 0.5% (annualized). The different possible expectations values will therefore be spaced by 0.5%, and go in both directions of the target. Assuming a higher precision of expectations formation (e.g. 0.01%) would not change our results qualitatively.[11] We will furthermore show that most of our results even carry over to a model with a continuum of possible expectation values (where there is no rounding and no discreteness in expectation formation).

Section 4.1 describes the specification of our benchmark model and shows that, as in the 3-type model, multiple steady states can exist. However, expectations are now required to be closer to perfectly self-fulfilling for this to occur. In Section 4.2 we show that the dynamics of our benchmark model in the absence of the zero lower bound are almost indistinguishable from those of a model where there is no discreteness in expectations, but, instead, a continuum of possible prediction values (the large type limit). Section 4.3 focuses on the zero lower bound on the nominal interest and the possibility of a self-fulfilling liquidity trap steady state in the benchmark model. Here we will let the anchoring of expectations vary endogenously with past deviations of inflation and output gap from their targets. We show that long lasting liquidity traps can arise. Whether the economy can recover from such a liquidity trap very much depends on how anchored expectations were when the economy fell in the trap, as well as on the inflation target.

4.1 Finitely many expectation values

4.1.1 Model

As in Section 3.1, the model is given by the law of motions for the economy ((4) and (5)), and by the equations describing expectations: (6), (7) and (8). However, instead of letting $H^x = H^\pi = 3$ in these expectation equations, we now allow H^x and H^π to differ

[10] Curtin (2010).

[11] Having expectations that are spaced closer together would make conditions for existence of multiple steady states more stringent. There would however continue to exist a range of policy parameters for which expectations are sufficiently self-fulfilling for multiple steady states to exist. In the limit where expectation values are infinitely close ot each other (the large type limit) this range of policy parameters is reduced to a single bifurcation point.

and to be considerably larger. The different biases that agents choose from ($b_{\pi,h}$ and $b_{x,h}$) will now be spaced by 0.5%, while extending equally to positive and negative values. For example, if $H^{\pi} = 41$, then the 41 possible biases for inflation will, in annualized values, be {-10, -9.5, -9,...,-0.5, 0, 0.5,..., 9, 9.5, 10}. Similarly, when $H^{x} = 35$, the output gap biases will range between 8% below target, and 8% above target (annualized). H^{x} and H^{π} therefore determine the largest deviations from target that agents consider, similar to the bias b in Section 3.1. We will therefore use H^{x} and H^{π} as measures of the *anchoring of expectations*.

Finally, we allow the fitness measures used in (6) to depend on the whole history of prediction errors, and not just the prediction error in the most recent period. That is,

$$U_{t-1}^{\pi,h} = -(1-\rho)(\pi_{t-1} - E_{t-2}^{h}\pi_{t-1})^2 + \rho U_{t-2}^{\pi,h}, \tag{16}$$

and

$$U_{t-1}^{x,h} = -(1-\rho)(x_{t-1} - E_{t-2}^{h}x_{t-1})^2 + \rho U_{t-2}^{x,h}. \tag{17}$$

4.1.2 Multiple steady states

We will first illustrate how almost self-fulfilling expectations can lead to the existence of multiple non-fundamental steady states where optimism or pessimism dominates, as in the 3-type model. We will do this for the case where $H^{x} = H^{\pi} = 41$, so that expectations are relatively unanchored and agents consider values up to ten percent above and below the inflation and output gap targets. In Section 4.3 we will look at how the zero lower bound on the nominal interest rate can create the possibility of a liquidity trap steady state in a specification with endogenously varying H^{x} and H^{π}.

Figure 2 plots a bifurcation diagram of a model with 41 types that is obtained by simulating the model for many different initial conditions for different values of the policy parameter ϕ_1. The intensity of choice is chosen relatively high ($\omega = 2 \cdot 10^6$) to facilitate coordination, and hence the existence of almost self-fulfilling non-fundamental steady states.

In the figure, it can be seen that for each of the 41 inflation prediction values there is a range of ϕ_1-values where this prediction comprises an *almost self-fulfilling steady state*. In particular, when ϕ_1 is around 0.95, then any value of inflation that agents could expect would lead to a realization of inflation that is close to this expectation value,

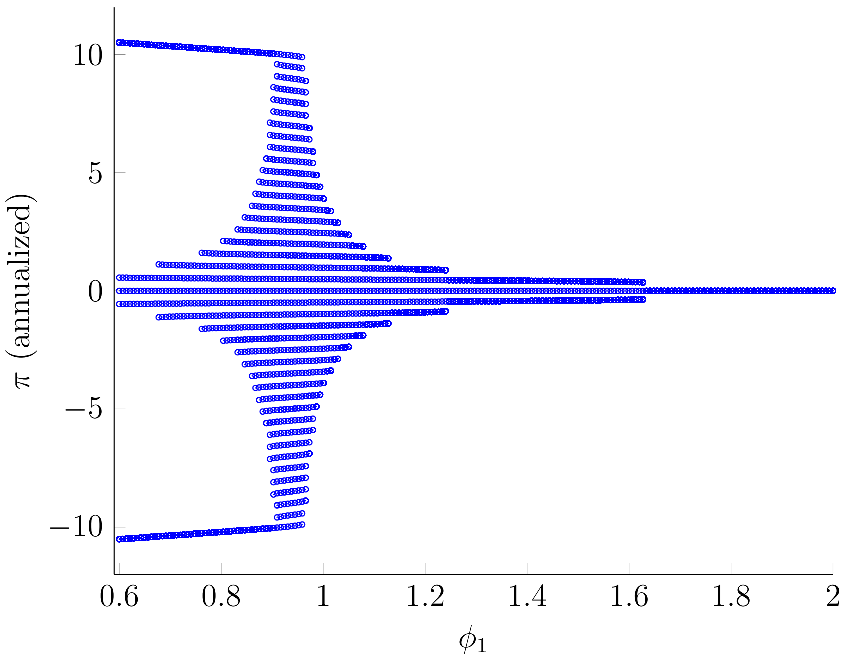

Figure 2: Bifurcation diagram of 41-type model in ϕ_1 for large intensity of choice ($\omega = 2 \cdot 10^6$). The blue lines represent (locally stable) steady states at the 41 different levels of inflation expectations.

so that agents will again choose these optimistic or pessimistic expectations. From a policy perspective this implies that for these relatively low values of ϕ_1, the central bank might have difficulty achieving its targets. Shocks to fundamentals of the economy (or to expectations themselves) could easily trigger *coordination on optimistic or pessimistic expectations* with corresponding high or low inflation and output gap.

When however the monetary authority responds more aggressively to inflation, non-fundamental expectations will lead to realizations that are less self-fulfilling due to stronger mean reversion to the central bank's targets. In the subsequent period agents will then adjust their expectations and become less optimistic or pessimistic. This leads to realizations of inflation even closer to the target, which again brings expectations closer to the fundamentals. As ϕ_1 increases, the non-fundamental steady states in Figure 2 therefore start to disappear. When ϕ_1 is larger than 1.3, only steady states very close to the central bank's targets exist, which might not be considered a problem from a policy perspective. As ϕ_1 is increased to 1.7 the fundamental steady state is unique and convergence to the targets will always occur in the long run.

When the central bank responds very weakly to inflation (e.g. $\phi_1 < 0.8$), then waves of optimism and pessimism become *self-reinforcing*. That is, low expectations of inflation lead to even lower realizations, which causes agents to become even more pessimistic.

This process continues until the lowest possible expectations are reached (lowest blue line in Figure 2). Analogously, self-reinforcing optimistic expectation lead to convergence to the highest possible expectation values (highest blue line in Figure 2).

For lower intensity of choice, the range of policy parameters for which intermediate steady state exist becomes smaller than in Figure 2. However, when monetary policy is too weak, it will always be the case that optimistic and pessimistic expectations are self-reinforcing, implying the existence of non-fundamental steady states where the most extreme expectation values dominate.

4.2 Large type limit

4.2.1 LTL model

We now consider the robustness of the findings of Section 4.1 to the assumption of finitely many discrete expectation values. In order to do this, we first combine equations (6), (7), and (16). Assuming memory $\rho = 0$ for analytical tractability, this can be written as

$$E_t \pi_{t+1} = \pi^T + \frac{\frac{1}{H} \sum_{h=1}^{H^\pi} b_h e^{-\omega(\pi_{t-1} - b_{\pi,h} - \pi^T)^2}}{\frac{1}{H^\pi} \sum_{h=1}^{H} e^{-\omega(\pi_{t-1} - b_{\pi,h} - \pi^T)^2}}. \tag{18}$$

Brock et al. (2005) show that if we let H^π go to **infinity** in this equation, the dynamics of our model can be closely approximated with the so called large type limit (LTL), where there is a **continuum** of prediction biases. Stated differently, the LTL, is an accurate description of the evolutionary selection, (18), of H^π forecasting rules when H^π is large. The LTL can be obtained by replacing sample means by population means in (18). Assuming that the prediction biases $b_{\pi,h}$ are normally distributed around zero with variance s^2, this gives

$$E_t \pi_{t+1} = \pi^T + \frac{\int_{-\infty}^{\infty} b e^{-\omega(\pi_{t-1} - b - \pi^T)^2} e^{-\frac{b^2}{2s^2}} \mathrm{d}b}{\int_{-\infty}^{\infty} e^{-\omega(\pi_{t-1} - b - \pi^T)^2} e^{-\frac{b^2}{2s^2}} \mathrm{d}b}. \tag{19}$$

The large type limit can also be interpreted in terms of Bayesian updating. Agents then try to learn in each period about the correct value of b, with $N(0, s^2)$ as their prior. The likelihood function (the distribution of π_{t-1} given the true value of b) is normal, with a variance inversely related to the intensity of choice (ω). This means that the intensity of choice (ω) is inversely related to the perceived noise with which agents can observe the correct value of b. This interpretation is in line with the random utility model underlying the multinomial logit probabilities given in (6) (see Anderson et al. (1992)). A more detailed comparison between the LTL and Bayesian updating is presented in Appendix B.1.

In the online supplementary material it is shown that (19) reduces to

$$E_t\pi_{t+1} = \frac{\pi^T}{2\omega s^2+1} + \frac{2\omega s^2}{2\omega s^2+1}\pi_{t-1}. \tag{20}$$

Analogously to the above, we obtain the following equation for output gap expectations.

$$E_t x_{t+1} = \frac{x^T}{2\omega s^2+1} + \frac{2\omega s^2}{2\omega s^2+1}x_{t-1}. \tag{21}$$

In the LTL model, expectations thus are a linear combination of the central bank's target, and of past realizations. The weights on these expectations depend on the intensity of choice (ω) and on the variance of the distribution of expectation values (s^2). Although the LTL allows agents to choose infinitely high and infinitely low expectation values, we can still talk about anchoring of expectations in this model. Where the anchoring of expectations was determined by b, and H^x and H^π in respectively the 3-type and the many type model, this role is taken by the *variance of the distribution of expectation values* (s^2) in the LTL model.

4.2.2 Comparison with 41-type specification

Figure 3 shows a simulated time series of inflation and output gap of the LTL model and the model with 41-types of Section 4. In the latter, we have set $\rho = 0$ in equations (16) and (17) for comparability with the LTL specification. We use the calibration of Woodford (1999) (with $\sigma = 0.157, \kappa = 0,024$ and $\beta = 0.99$), and further set annualized s^2 equal to 0.35, so that the variance of the distribution of types in the LTL model matches that of the 41-type model. We calibrate the intensity of choice at $\omega = 63500$ to let the 41-type model match expectations from survey data.[12] At this calibration, the interquartile range of the expectations distribution in the fundamental steady state is 1% (in annualized terms). Outside of this steady state the interquartile range then typically is 1.5%, and less for realized values close to the highest or lowest possible prediction value.

[12]Note that the magnitude of the intensity of choice depends on the units of measurement of the data. 63500 should therefore not necessarily be seen as a high number. If we interpret the LTL model in terms of Bayesian updating, then $\omega = 63500$ implies that the perceived noise agents encounter in observing true values has a standard deviation of 1.12% of annualized inflation (see Appendix B.1).

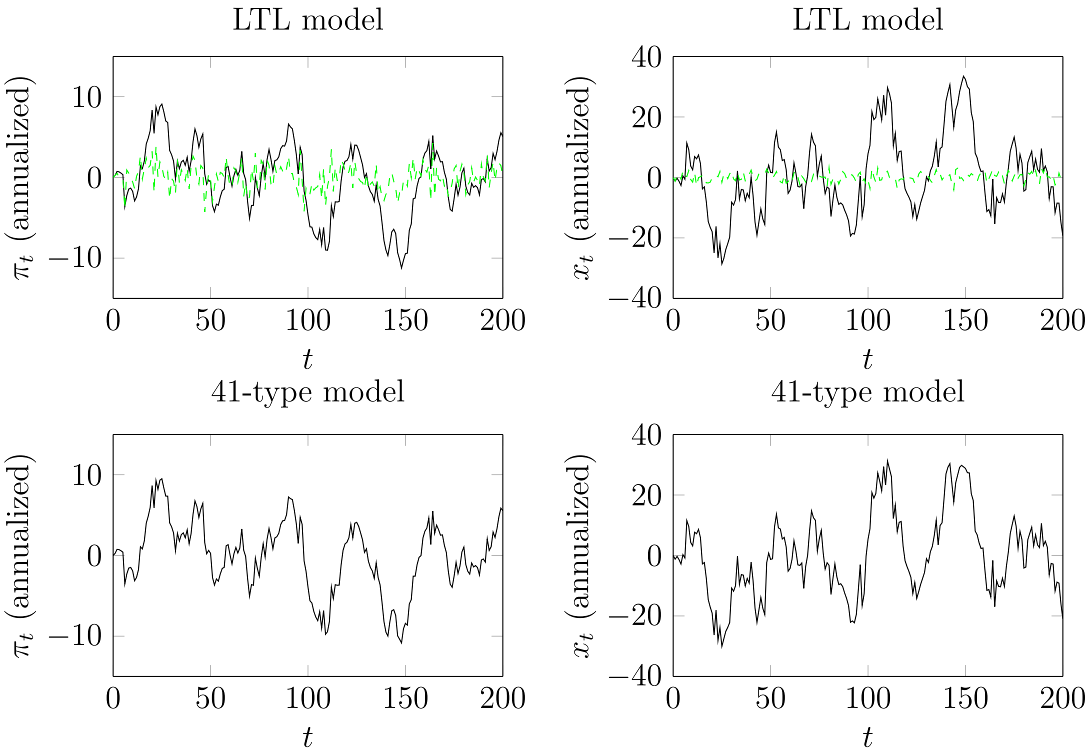

Figure 3: Simulated time series for LTL model (top panels, black) and 41-type model (bottom panels). The case of rational expectations is plotted in the top two panels in dashed green.

This is in line with the findings of Mankiw et al. (2003), who show that the interquartile range of the Livingston Survey and the Survey of Professional Forecasters is around 1%.

Shocks to inflation and output gap are white noise, and have an annualized standard deviation of 1.5%. In the interest rate rule we set $\pi^T = 0$, $\phi_1 = 1.5$ and $\phi_2 = \sigma = 0.157$.

The first thing to notice in Figure 3, is that the time series of both the LTL model and the 41-type model are highly **persistent**. Since there is no autocorrelation in shocks, this persistence is fully endogenous, and comes from the way agents form expectations. Positive or negative shocks can initiate *waves of optimism or pessimism* that are eventually reversed. If agents would form expectations rationally (which, in our model with white noise shocks, would mean that they always expect the targets) such drifts in inflation and output gap would not have arisen. For comparison, the case of rational expectations is plotted in dashed green in the upper two panels.

The second thing to note from Figure 3, is that the time series of the LTL model and the 41-type model look **almost identical**. So, even though in the 41-type model expectations are discrete and very high or low values of inflation and output are not considered when agents form expectations, the resulting dynamics are here practically the same as in the LTL model with a continuum of expectations values ranging from minus infinity to infinity.

The intuition for both the waves of optimism and pessimism, and for the striking similarities between the two model specifications, are (almost) self-fulfilling expectations. When shocks drive up inflation, this leads agents to increase their inflation expectations. Since the main determinant of inflation is expected inflation, inflation is likely to be high again in the subsequent period, leading again to high expectations. Meanwhile, the monetary authority responds to the high inflation expectations by increasing the interest rate, which depresses output gap. This leads to more pessimistic expectations about output gap, amplifying the recession.

For the LTL model, these almost self-fulfilling expectations imply very slow convergence to the steady state and a **near unit root**.[13] That is, in the absence of further shocks, realizations of inflation would be closer to the target then in the previous period, but not by much. Therefore, expectations would also only slowly adjust and mean reversion back to the steady state would go quite slowly. When new shock innovations are added to this near unit root process, the near random walks of Figure 3 arise, with waves of optimism and pessimism.

We saw in Section 4.1.2 that in the 41-type model almost self-fulfilling expectations can lead to the existence of multiple steady states for high intensity of choice, due to coordination on a specific expectation value. However, even when the intensity of choice is lower and monetary policy is aggressive enough to make sure that expectations are not sufficiently self-fulfilling for this to occur, it might still be the case that expectations are self-fulfilling enough to induce agents to *only slowly* switch to another expectation value. This is what is happening in the bottom panels of Figure 3.

The waves of optimism and pessimism described above are somewhat limited by the anchoring of expectations. In the bottom panels of Figure 3 (41-type) it regularly occurs that the lowest or highest possible expectation values have the best fitness. If expectations were less anchored, this would induce some agents to have even lower or even higher expectations in these cases. In the top panels of Figure 3 (LTL), agents can always choose even higher or lower expectations. However, the variance of the distribution of expectations types, which determines the anchoring of expectations in this model, still

[13]We show that the model has a near-unit root for this calibration in Section 5.1

keeps most of the expectation mass relatively close to the target. That is, in both models the anchoring of expectations prevents waves of optimism and pessimism from becoming too large.

We can conclude that when monetary policy is active, the 41-type model gives very similar expectations and dynamics as the LTL model, even though expectations are here assumed to take on only discrete values, and even though the 41-type model places an upper and lower bound on expectation values that agents can choose from. In Sections 4.3 and 5.2 we will see that in a liquidity trap with a binding zero lower bound, the existence or absence of a lower bound on possible expectation values does qualitatively alter dynamics.

4.3 Liquidity traps in many type model

4.3.1 Model with endogenously varying anchoring of expectations

In sections 4.2.2 we saw that self-fulfilling waves of optimism and pessimism can arise in our model, even in the absence of the zero lower bound on the nominal interest rate. We now return to our main question, and investigate whether our model can explain self-fulfilling liquidity traps.

We do this with the model with finitely many expectation values of section 4.1 in its most general form. That is, expectations are given by (6), (7) and (8) and we allow the anchoring of expectations to differ between inflation (H^{π}) and output gap (H^{x}) and to **vary over time**. When expectations are such that the nominal interest rate is positive, the model is in the "positive interest rate region", and the economy evolves according to (4) and (5). When the zero lower bound on the nominal interest rate is binding, the model is in the "ZLB region", and inflation and output gap are given by (11) and (12).

When the model is in the ZLB region, there is no feedback on expectations through monetary policy, and expectations can easily become self-reinforcing. In Section 3.2 we saw that in the 3-type model this can lead to the existence of a liquidity trap steady state with coordination on pessimistic expectations. We will now investigate whether *coordination on a liquidity trap steady state with self-reinforcing expectations* can also arise in our many type model.

Before we can address this question, we need to specify a law of motion for the anchoring of expectations. We do this by assuming that expectations become *less* anchored when realizations are far away from the target for an extended period of time, and become *more* anchored when realizations continue to be relatively close to the central bank's targets. We model this by letting agents adaptively update the largest absolute deviation from target (LAD) that they consider, based on new realizations of absolute deviations of variables from target. That is,

$$LAD^z_t = \gamma LAD^z_{t-1} + (1-\gamma)|z_{t-1} - z^T|, \qquad z = \pi, x. \tag{22}$$

The time t anchoring of expectations, H^z_t, is obtained from annualized LAD^z_t by first multiplying by 4, and then truncating to the lowest odd integer.[14] γ is a persistence parameter that determines how fast expectations can become more or less anchored.

Note that with the anchoring of expectations evolving according to (22), expectations would eventually become **fully anchored** in the target steady state. When the model is subject to shocks however, the economy never fully reaches this steady state, but instead inflation and output gap keep fluctuating around their targets. In this case, expectations never become fully anchored, and there will *always* be heterogeneity in expectations.[15]

4.3.2 Simulation with different initial anchoring of expectations

We now simulate the above model for different levels of the initial anchoring of expectations. With this simulation exercise we can investigate whether liquidity traps can arise, and how this depends on the anchoring of expectations. Figure 4 shows time series of inflation and output gap (top two panels) for two different values of initial anchoring. H^π_t and H^x_t evolve over time and are plotted in the middle two panels. Green curves depict the case where the anchoring of expectations start out at $H^\pi_0 = H^x_0 = 41$, while the purple curves depict the case of initial anchoring of $H^\pi_0 = H^x_0 = 33$. That is, either agents initially consider expectation values up to 10% above and 10% below the target, or expectations are somewhat more anchored, and agents only consider values between

[14]It is multiplied by 4 because an extra percentage point in LAD^z_t implies a percentage point in the positive and a percentage point in the negative direction, and hence 4 steps of 0.5%. Truncating to the nearest *odd* integer ensures that symmetry in expectation values is maintained.

[15]Also note that the fractions of agents expecting different expectation values evolve according to a probability distribution, where in any period each agent has a chance of choosing a particular value. It is therefore not necessarily the case that there are individual agents who are either constantly too optimistic or constantly too pessimistic and hence make persistent forecast errors. Instead, some agents will be too optimistic for a while and then too pessimistic, while other agents are first too pessimistic and then too optimistic.

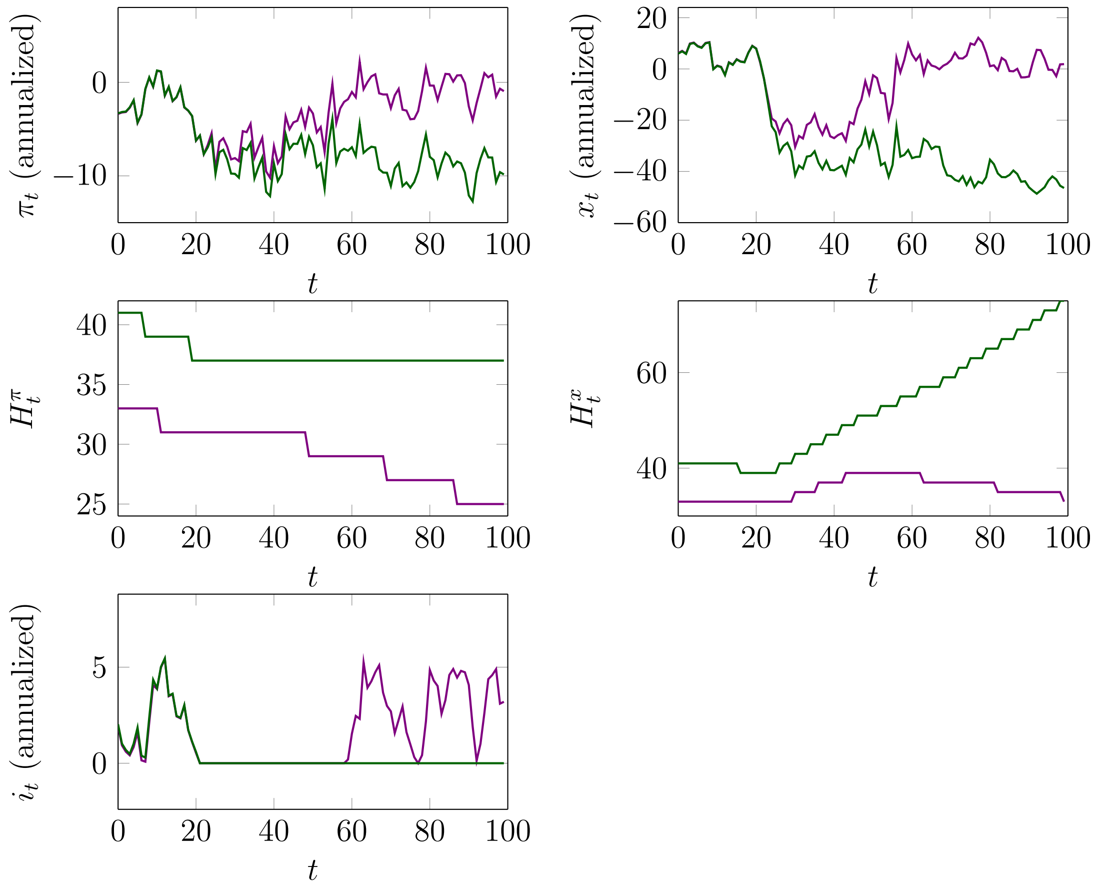

Figure 4: Simulated time series of model with endogenous anchoring of expectations and ZLB. The purple and green curves represent time series with different initial anchoring of expectations H_0^π and H_0^x.

−8% and 8% around the target. We deliberately take two initial conditions that are relatively close to each other, in order to illustrate that a small difference in initial anchoring of expectations can have a large impact on subsequent dynamics. Finally, the nominal interest rate is plotted in the bottom panel of Figure 4.

In the figure, we have used the calibration of Section 4.2.2, but we allow the fitness measures, (16) and (17), to depend on the whole history of realizations by setting $\rho = 0.5$. We furthermore set $\gamma = 0.995$ so that expectations can become more or less anchored endogenously, but that this happens at a much slower rate than the switching of agents between expectation values. This reflects the intuitive assumption that expectations remain relatively anchored when shocks drive inflation and output gap far away from their targets for only a couple of periods, but that the range of expectation values that agents consider to be realistic starts increasing as variables significantly deviate from their targets for *longer* periods of time.

In Figure 4, it can be seen that the economy starts out with low inflation and high output gap.[16] Subsequently, both variables revert back to their target and stay close to it for a while. This induces inflation expectations to become *more anchored* (LAD_t^π and H_t^π decrease), as can be seen in the middle left panel.

Around period 12, some bad shocks hit the economy, which drive down inflation. Almost self-fulfilling expectations then induce a wave of pessimism about inflation and optimism about output gap, as also sometimes happens in Figure 3. However, at some point (after period 20) the low inflation leads to a binding zero lower bound, and the nominal interest rate is set to zero. As inflation falls even more, the *real interest rate* becomes higher and higher, leading output gap to start falling as well. Since there is no feedback from monetary policy anymore, pessimistic expectations now are self-reinforcing, just as in the dark shaded area in Figure 1, in the 3-type model.

Inflation and output gap keep falling until output gap expectations start to hit their lowest possible value (Around period 30). The economy now is stuck in a *liquidity trap steady state*, both in case of the purple and of the green curve. In the case of less anchored expectations (the green curves), output gap expectations can fall more however, leading to lower realizations of output gap and inflation than in the case of more anchored initial expectations (purple curves). As the liquidity trap with low inflation and output gap continues to last, output expectations start becoming less anchored (LAD_t^x and H_t^x increase), as can be seen in the middle right panel of Figure 4.

Then, around period 45, favorable shocks start to hit the economy, and this is where the two curves really start to differ. In the case of *more* anchored expectations (purple curves), where inflation and output gap have not become too low yet, these shocks are enough to get the economy out of the liquidity trap. Inflation and output and their expectations now become high enough for the model to revert back to the positive interest rate region, from which monetary policy can achieve reversion back to the target. As this happens, expectations become more and more anchored (middle panels).

In case of *less* anchored expectations (green curves) the favorable shocks also increase inflation and output gap. However, here the liquidity trap already was too deep, so that the zero lower bound remains binding, and both variables quickly start to fall again. During the remaining periods, the economy keeps fluctuation around the liquidity trap

[16] We first run our model for 50 periods to initialize it (this is necessary for the fitness measures). The initial levels of inflation and output gap are therefore random.

steady state, which has lower and lower inflation and output gap values as output gap expectations keep becoming more unanchored (middle right panel). It therefore becomes more and more difficult for favorable shocks, or policy interventions, to get the economy out of this liquidity trap.

The key policy insight that can be obtained from this simulation exercise is the following. Since the initial anchoring of expectations at the time bad shocks hit the economy is crucial for eventual recovery of a liquidity trap, the central bank must make sure that this **initial anchoring is strong enough**. It can do this by making sure that in the years before the bad shocks hit, inflation and output gap are stabilized close to their targets. Since the zero lower bound was not yet a binding constraint during these periods, this can be achieved with adequate conventional monetary policy. In Section 4.1 and 4.2 we showed that (without a binding zero lower bound) the existence of multiple steady states in our benchmark model due to almost self-fulfilling expectations, can be prevented with aggressive enough monetary policy. In Section 5.1 we will give analytical conditions for local stability and eigenvalues in the large type limit model to provide robustness of this result, and to give more insight in how conventional monetary policy can exclude almost self-fulfilling expectations.

First, we will however stick with our benchmark model under the zero lower bound, and show that a stronger initial anchoring of expectations not only makes recovery more likely, but can even exclude the possibility of a liquidity trap altogether.

4.3.3 Liquidity traps and anchoring of expectations

To gain more insight in the role of the anchoring of expectations, and to provide robustness to the above simulation exercise, we now turn to a bifurcation analysis. Figure 5 plots the steady state values of both the target steady state (upper blue dots) and the liquidity trap steady state (lower blue dots), for different levels of H_t^π and H_t^x. Additionally, we plot the unstable steady state (green dots), lying between these two steady states. This unstable steady state separates the **basin of attraction** of the target steady state from that of the liquidity trap steady state.[17] That is, in the absence of shocks, inflation and output gap on one side of the unstable steady state imply convergence to the target steady state, while initial inflation and output gap on the other side of the unstable steady state

[17]More precisely, the unstable steady state is a saddle point, and the basins of attraction of the two stable steady states are separated by the stable manifold of the saddle point steady state.

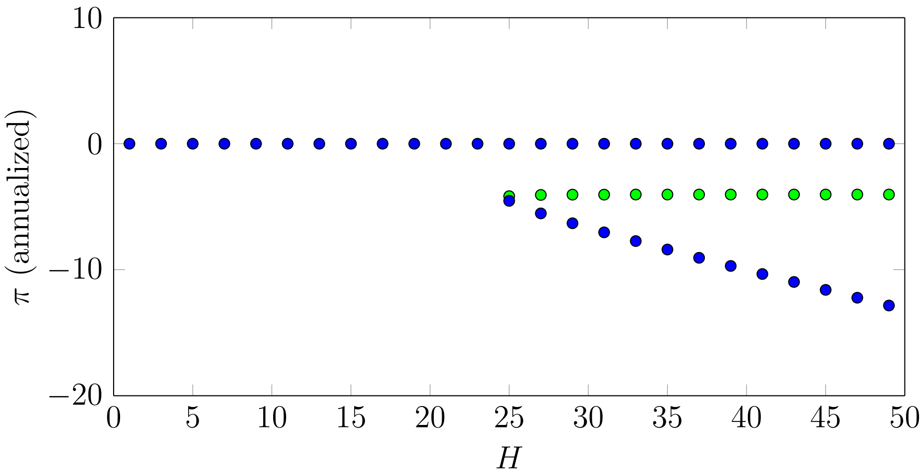

Figure 5: Stable target steady state (top blue dots); stable liquidity trap steady state (bottom blue dots), for different levels of anchoring of expectations in the model with finite expectation types. The unstable steady state in the middle (green dots) separates their basins of attraction.

imply convergence to the liquidity trap steady sate.

To keep the graphical representation clear, we do not show the output gap value of the steady state, and put its inflation value on the vertical axis. Additionally, we vary H_t^π and H_t^x together and show on the horizontal axis $H = H_t^\pi = H_t^x$. This allows us to keep the plot 2-dimensional, without loss of generality.

It can be seen in Figure 5 that for strong anchoring of expectations (low H) the target steady state is unique, and a persistent liquidity trap can not arise. However, as expectations become less anchored, the liquidity trap steady state comes into existence ($H = 25$). It then however still lies relatively close to the target steady state, and, more importantly, very close to the unstable steady state. This implies that small favorable shocks can bring inflation out of the liquidity trap, and above the unstable steady state, from which the system will converge to the target steady state. As expectations become even more unanchored, the liquidity trap steady state has a lower inflation value and lies farther away from the unstable steady state. Recovery due to favorable shocks is less likely in this case. This is exactly what we saw in Figure 4. In the purple time series, inflation was still close to the basin of attraction of the target steady state (i.e. close to the unstable steady state), and favorable shocks led to recovery. In the green time series, inflation was too far away from the unstable steady state, and favorable shocks did not bring the economy back to the basin of attraction of the target steady state.

The unstable steady state in Figure 5 lies very close to the negative of the steady state real interest rate: $-\bar{r}$. We will show in Section 5.2 that this is no coincidence and that for the LTL model with a continuum of expectation values, the unstable ZLB steady state converges exactly to this value. The intuition for this result is that the (log-linearized) period t real interest rate (given by $i_t - \bar{r} - E_t\pi_{t+1}$) is positive under a binding zero lower bound if and only if $E_t\pi_{t+1} < -\bar{r}$. It follows from Equation (1), that when this is the case, output gap realizations will be lower than output gap expectations. It then follows from (2) that for reasonable calibrations of κ and β the same holds for inflation. This implies that only if inflation expectations are low enough compared to the steady state real interest rate, pessimistic expectations will be *self-reinforcing* and hence lead to convergence to the liquidity trap steady state. A necessary condition for inflation expectations to become low enough is that H is large enough, so that the lowest possible inflation values that agents consider is lower than $-\bar{r}$. If expectations are more anchored (lower H) and agents do not consider inflation values below $-\bar{r}$, then $E_t\pi_{t+1} < -\bar{r}$ can never hold and the liquidity trap steady state *does not exist*. Hence, if the central bank manages to anchor expectations strongly enough to its targets (by conducting aggressive enough monetary policy), it can *exclude* to possibility of bad shocks driving the economy into a self-reinforcing liquidity trap.

4.3.4 Increasing the inflation target

Since agent's expectations are distributed around the targets, another way the central bank could reduce the likelihood (or even exclude the possibility) of a self-reinforcing liquidity trap, is by choosing a higher inflation target. For a given anchoring of expectations, the *most pessimistic expectations* that agents consider would then be higher. This would make it less likely that $E_t\pi_{t+1} < -\bar{r}$ will hold, and that the economy enters a liquidity trap.

In Figure 6 we plot the target steady state (top blue curve) and the liquidity trap steady state (bottom blue curve) as a function of the inflation target (π^T), for the case of $H = 41$. Note that, as the inflation target is chosen higher, the target steady state has an equivalently higher inflation level as well, making the top blue curve upward sloping. Since agents consider 10% below the inflation target as their lowest possible value, the bottom blue line is also upward sloping. Figure 6 also plots the unstable steady state (green curve) separating the basin of attraction of the liquidity trap from that of the

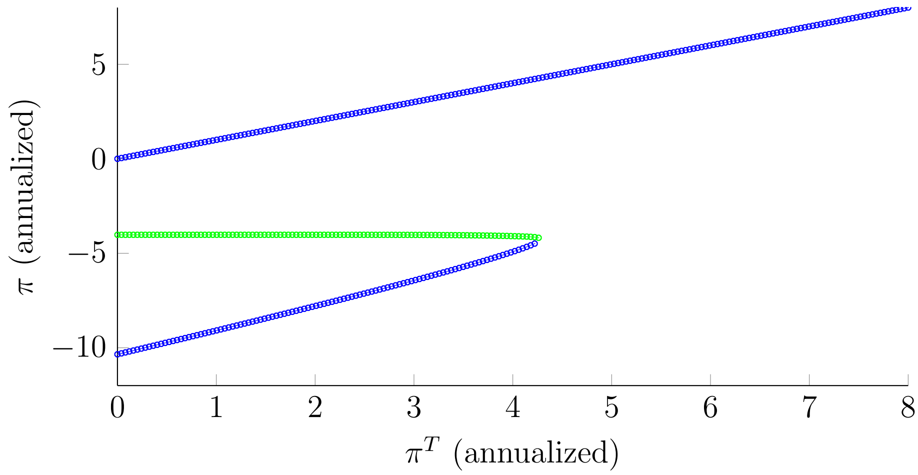

Figure 6: Bifurcation diagram of 41-type model in π^T. The upper blue curve represents the fundamental steady state, and the lower blue curve the liquidity trap steady state. The green curve depicts the unstable steady state that separates their basins of attraction.

target steady state.

In Figure 6, it can be seen that for a zero inflation target, the liquidity trap steady state is associated with a quite low level of inflation and lies far away from the basin of attraction of the target steady state. As the inflation target is increased, the liquidity trap steady state comes closer to the unstable steady state (green curve), making it more likely that shocks bring the economy from the liquidity trap steady state to the basin of attraction of the target steady state. As the inflation target is increased even more, the liquidity trap steady state disappears altogether, and convergence to the target steady state *always* occurs. Note however, that if we would let H vary endogenously as in Figure 4, a high inflation target can only temporarily exclude the possibility of a liquidity trap. If, for some reason (e.g. bad shocks), inflation realizations remain low for a long period of time, expectations would become *less anchored*, and agents might consider expectation values that satisfy $E_t\pi_{t+1} < -\bar{r}$, even when the inflation target is high. In that case, the liquidity trap steady state would exist again.

The above result is similar to findings under rational expectations, where liquidity traps can be prevented by choosing the inflation target high enough, so that large shocks in the model do not lead to a severely binding zero lower bound and a high real interest rate.[18] Our model however emphasizes the role of expectations rather than that of exogenous shocks. Low expectations might come about because of large shocks, but could also

[18]See e.g. Ball (2013)

arise slowly because of declining confidence after a period where small negative shocks were relatively more prevalent.

5 Analytical results of LTL model

In the previous section we showed that the initial anchoring of expectations when the economy hits the zero lower bound is crucial in determining whether recovery can occur. We concluded that during normal times the monetary authority would do well to actively try to **anchor expectations**, and that it could do so by preventing (almost) self-fulfilling expectations and waves of optimism and pessimism. In Section 4.2 we showed that, without the zero lower bound, the dynamics of our benchmark model are almost indistinguishable from those of the large type limit model where agents can choose from a continuum of expectation values. Therefore, we can shed more light on how monetary policy can prevent (almost) self-fulfilling expectations by studying the eigenvalues of the analytically tractable LTL model. We do so in Section 5.1. In Section 5.2 we study the LTL model under the zero lower bound to provide robustness to the results of Section 4.3, and to give more insight in the occurrence of liquidity traps.

5.1 Stability in LTL model without ZLB

In the LTL model, a locally stable target steady state (all eigenvalues inside the unit circle) implies that convergence to the targets can occur. This is a first requirement for inflation and output gap to remain close to their targets and hence for expectations to become more anchored in a model with endogenous anchoring as presented in Section 4.3. However, if one eigenvalue lies only just inside the unit circle (a *near unit root*), then convergence to the steady state will be very **slow** and expectations will not easily become more anchored. Below, we therefore first provide conditions for local stability and then focus on the occurrence of near unit roots.

Proposition 5 states the conditions for stability of the target steady state. There are two conditions, which respectively ensure that the first eigenvalue of the dynamical system, λ_1, is smaller than $+1$, and that the other eigenvalue, λ_2, is larger than -1.[19]

[19]Note that eigenvalues can also be complex, but that they always lie inside the unit circle when this is the case.

Proof of Proposition 5 is given in Appendix B.2.[20]

Proposition 5 (**Stability LTL**). *(See Figure 7) In the LTL model, the target steady state steady state is locally stable if and only if*

$$\phi_1 > 1 - \frac{1+(1-\beta)2\omega s^2}{2\omega s^2+1}\left(\frac{\sigma}{2\omega s^2\kappa} + \frac{\phi_2}{\kappa}\right), \quad (\lambda_1 < +1), \tag{23}$$

and

$$\phi_1 < 1 + \frac{(1+\beta)2\omega s^2+1}{2\omega s^2+1}\left(\frac{(4\omega s^2+1)\sigma}{2\omega s^2\kappa} - \frac{\phi_2}{\kappa}\right), \quad (\lambda_2 > -1). \tag{24}$$

It follows from Proposition 5 that there is both a lower and an upper bound on how aggressive monetary policy responses can be. The conditions are presented in terms of the inflation policy coefficient, ϕ_1, and are a function of the output gap policy coefficient, ϕ_2, the anchoring of expectations s^2 (see Subsection 4.2.1), and other model parameters. The upper bound does not occur for realistic (policy) parameter settings, and it is also not very intuitive that instability due too aggressive monetary police would occur in practice.[21] We therefore focus on Condition (23).

Figure 7 plots the stability condition (23) as a function of s^2 and ϕ_2. In this figure we used the same calibration for σ, κ, β and ω as in Section 4.2.2. In the figure, it can be seen that for low values of s^2 (where expectations are strongly anchored to the targets), weak inflation policy ($0 < \phi_1 < 1$) does not lead to instability. For higher values of s^2 (unanchored expectations) however, the central bank must respond strongly enough to inflation in order to satisfy (23). If we let s^2 go to infinity, Condition (23) reduces to $\phi_1 > 1-(1-\beta)\frac{\phi_2}{\kappa}$, which is the well known *Taylor principle* that must be satisfied under rational expectations in order to obtain local determinacy.[22]

Even though Condition (23) can be satisfied with relatively weak monetary policy, we saw in Figure 3 that even when $\phi_1 = 1.5$ and $\phi_2 = 0.157$ white noise shocks can trigger

[20] A similar proposition for a contemporaneous interest rate rule can be found in Pecora and Spelta (2017).

[21] If monetary policy is so aggressive that Condition (24) is violated, the following occurs. If the CB responds too strongly, high inflation expectations lead too very low inflation realizations, which (depending on how strongly expectations are anchored), can lead to very low expectations in the subsequent period. These low expectations then again induce a strong policy response that leads to very high realizations of inflation, and high expectations in the subsequent period. The explosive overshooting then continues with both expectations and realizations reaching ever higher and lower values.

[22] The result that stability is more easily achieved for low values of s^2, is similar to findings in models with constant gain learning. Evans et al. (2009) for example find that stability conditions become less stringent as the gain parameter goes to zero. Where, however, a gain parameter of zero implies full anchoring to *initial beliefs*, in our model, $s^2 = 0$ implies full anchoring to the *targets of the central bank*.

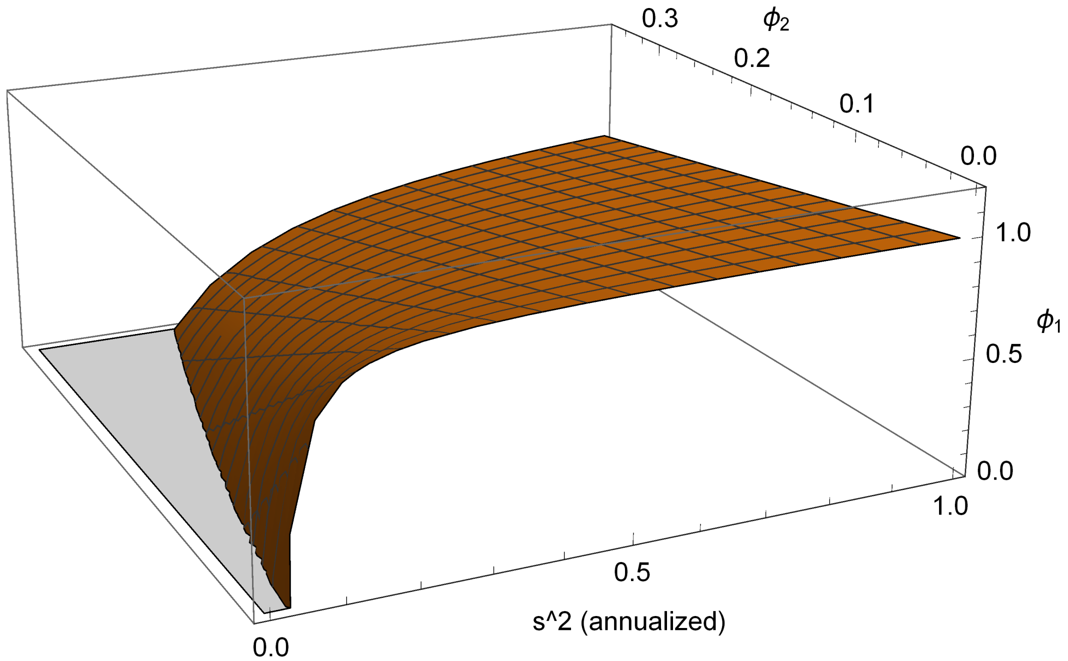

Figure 7: Stability region of the Target steady state in the LTL model for different policy parameters (ϕ_1, ϕ_2) and anchoring of expectations (s^2). The target steady state is locally stable above the plotted surface.

persistent waves of optimism and pessimism, due to almost self-fulfilling expectations. The largest eigenvalue in the LTL model is then smaller than 1 in absolute value (so that local stability is achieved), but still *close* to 1. Under such a near unit root, dynamics become very *persistent* and mean reversion to the steady state is slow.

Figure 8 plots the largest eigenvalue of the LTL model as a function of ϕ_2, for different values of ϕ_1. Here, we use again the calibration of Section 4.2.2, also for s^2.[23] The figure illustrates that when ϕ_2 is relatively low, a near unit root (absolute value of eigenvalue close to 1) occurs, no matter how aggressive the monetary authority responds to inflation (high ϕ_1). However, when ϕ_2 is chosen larger, a near unit root can still occur if ϕ_1 is not large enough, or if ϕ_2 is chosen too large. We can conclude that a *well tuned combination* of ϕ_1 and ϕ_2 is required to prevent nearly self-fulfilling waves of optimism and pessimism.

5.2 LTL under the Zero Lower Bound

In Section 4.2.2 we saw that without the zero lower bound, the LTL model and 41-type model share striking similarities (Figure 3). We now investigate whether the policy

[23]For higher values of s^2 all curves are shifted upward, so that the largest eigenvalue becomes higher for a given combination of policy parameters, and for lower values of s^2 all curves of Figure 8 are shifted downward.

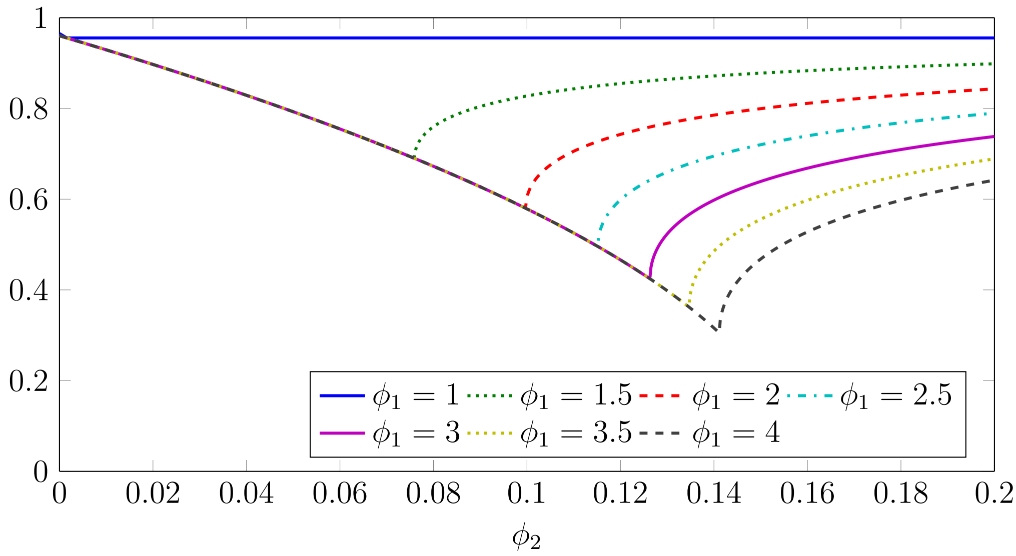

Figure 8: Absolute value of largest eigenvalue of LTL model as function of ϕ_2 for different values of ϕ_1.

implications under the zero lower bound of Section 4.3 also carry over to the LTL model. Since we can obtain analytical results for the LTL model, the analysis of this section can shed more light on the occurrence of liquidity traps.

As in the model with finitely many expectation values, the possibility of a persistent liquidity trap in the LTL model depends on the existence and position of an unstable steady state that limits the basin of attraction of the target steady state (the green dots/line in Figures 5 and 6). However, unlike the finite expectation model, initial conditions outside the basin of the target steady state do not lead to convergence to a liquidity trap steady state in the LTL model. Since no lower bound on expectations exists in this model, self-reinforcing pessimistic expectations will lead agents to *keep adjusting their expectations downward*. This gives similar dynamics to the green time series in Figure 4, where expectations keep decreasing as they become less anchored. However, in the LTL model expectations can fall faster, bringing the economy quickly into a *deflationary spiral* with ever decreasing inflation and output gap. Such a deflationary spiral under the zero lower bound is, amongst others, also found by Evans et al. (2008) and Benhabib et al. (2014), under adaptive learning.

Proposition 6 states the inflation and output level of this unstable saddle point, and its condition for existence. Finally, the line in the inflation output gap plane is given that goes through the unstable steady sate, and is the boundary of the basin of attraction of the target steady state. Proof of the proposition is given in Appendix B.3.

Proposition 6 (ZLB dynamics LTL). *The unique ZLB steady state of the LTL model is given by*

$$\pi^{zlb} = \frac{\left(\sigma(1+(1-\beta)2\omega s^2) + \kappa(2\omega s^2+1)\right)\pi^T + \kappa(2\omega s^2+1)^2\bar{r}}{\sigma(1+(1-\beta)2\omega s^2) - \kappa(2\omega s^2+1)2\omega s^2}, \tag{25}$$

$$x^{zlb} = \frac{\left((1-\beta)\frac{\sigma}{\kappa}(1+(1-\beta)2\omega s^2) + (2\omega s^2+1)\right)\pi^T + (1+(1-\beta)2\omega s^2)(2\omega s^2+1)\bar{r}}{\sigma(1+(1-\beta)2\omega s^2) - \kappa(2\omega s^2+1)2\omega s^2}. \tag{26}$$

This steady state is an unstable saddle-point, and exists if and only if

$$\omega s^2 > \frac{1}{4}\left(\sqrt{(1-(1-\beta)\frac{\sigma}{\kappa})^2 + 4\frac{\sigma}{\kappa}} - (1-(1-\beta)\frac{\sigma}{\kappa})\right). \tag{27}$$

When the steady state exists, all initial conditions above the stable eigenvector through (π^{zlb}, x^{zlb}) *imply recovery to the positive interest rate region, while all initial conditions below it result in a deflationary spiral. This eigenvector has slope*

$$-\frac{1-(1-\beta)\frac{\sigma}{\kappa} + \sqrt{(1-(1-\beta)\frac{\sigma}{\kappa})^2 + 4\frac{\sigma}{\kappa}}}{2\sigma}. \tag{28}$$

Figure 9 illustrates the "recovery region" and the "deflationary spiral region" in the (π, x)-plane. Here we set $\pi^T = 2\%$, and otherwise use the calibration of Section 4.2.2. The red line in Figure 9 depicts the zero lower bound. For values of inflation and output gap above this line, the nominal interest rate is positive (positive interest rate region) and convergence to the target steady state can occur. For combinations of inflation and output gap below this line, the zero lower bound is a binding constraint (ZLB region). The black line indicates the stable eigenvector through the unstable ZLB saddle steady state of Proposition 6. Combinations of inflation and output gap above (i.e. to the right of) this line are not too low, so that recovery to the positive interest region occurs. Inflation and output gap combinations below the black line lead to declining inflation and output gap and hence a deflationary spiral.[24]

The position of the black line, and hence the size of the recovery region, depends on the position of the ZLB steady state, which in turn depends on the anchoring of

[24] Our "recovery region" is strongly related to the "corridor of stability" discussed in Benhabib et al. (2014) and Eusepi (2010). These papers consider global dynamics of a New Keynesian model under adaptive learning, and find numerically that there is a region around the ZLB steady state from which the economy goes to the target steady state and a region from which inflation and output go to minus infinity. In our log-linearized model we are able to find an analytical expression for the curve that separates these two regions. The drawback of using a linearized model is of course that it only gives reasonable approximations for inflation and output gap not too far from the steady state.

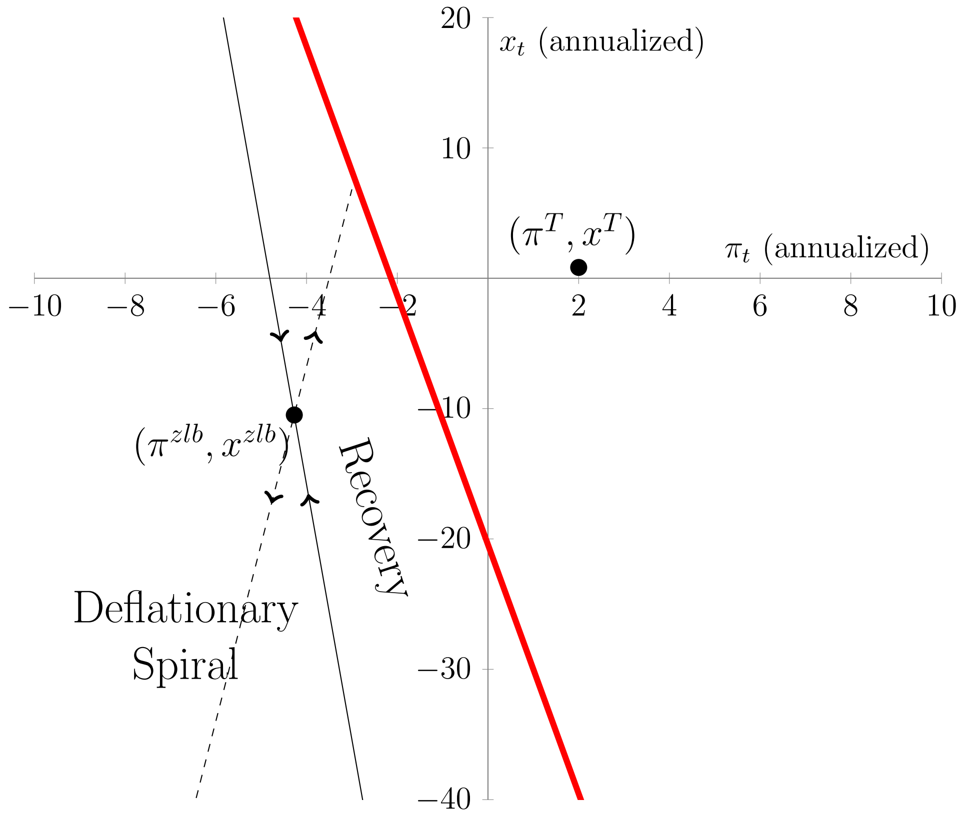

Figure 9: Regions of recovery and a deflationary spiral in the annualized (π, x)-plane. The thick red line indicates the ZLB. The black dot at 2% inflation indicates the target steady state, while the black dot at the bottom of the figure depicts the unstable ZLB saddle steady state from Proposition 6. The unstable eigenvector through this saddle is depicted by the dashed line, while the stable eigenvector is depicted by the solid black line. For initial conditions to the left of this black line a deflationary spiral occurs, and for initial conditions to the right of this line inflation and output gap will recover and increase to the positive interest rate region.

expectations and the inflation target. Figure 10 shows how exactly the position of the ZLB steady state depends on the anchoring of expectations, with s^2 on the horizontal axis and inflation on the vertical axis. For comparability with Figure 5 we have set the inflation target back to zero in the figure. The dashed curve represents the unstable ZLB steady state, while the solid line depicts the stable target steady state.

Taking the limit $s^2 \to \infty$ in (25) gives (for any value of the inflation target) $-r$. Therefore, the green line in Figure 10 must have a *horizontal asymptote* at $-r$. As commented earlier, the position of the unstable steady state in the model with finite expectation values lies very close to this value as well. This is confirmed by comparing the green curves in Figure 5 and Figure 10.

However, the way the unstable steady state seizes to exist for more anchored expecta-

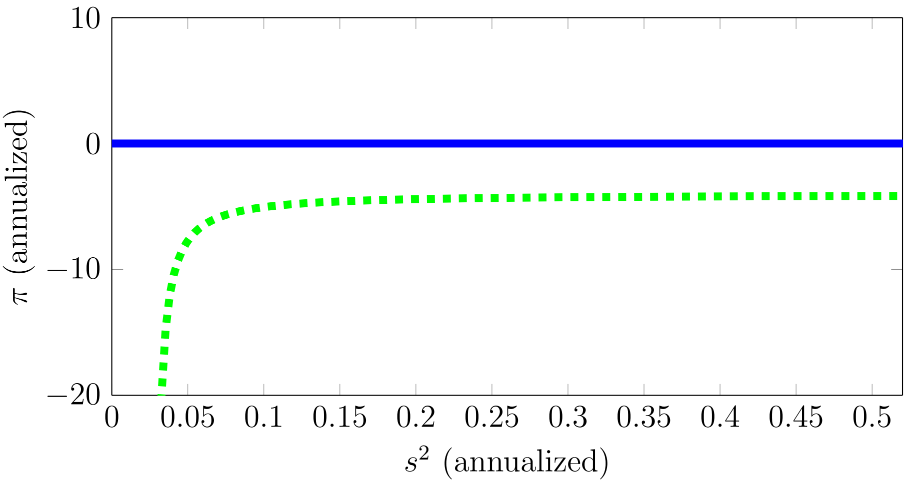

Figure 10: Stable target steady state (solid blue) and ZLB saddle steady state (dashed green) In LTL model for different levels of anchoring of expectations. Points above the ZLB saddle steady state imply convergence to the target steady state, while points below it lead to a deflationary spiral.

tions is different across the two models. In Figure 5 the unstable steady state disappears when expectations are so strongly anchored that the lowest expectation values that agents consider no longer lead to low enough inflation and output gap for expectations to be self-reinforcing. This happens when H is between 23 and 25, which would correspond to an anchoring of expectations in the LTL model of around $s^2 = 0.12$ (annualized). In Figure 10 we see that for (annualized) $s^2 < 0.12$, the steady state continuous to exist, but that it lies at a lower and lower inflation value. This reflects that even though in this model it is always possible that agents choose lower and lower expectation values, the probability for this to happen becomes smaller as the variance of the distribution of expectation values decreases. Realizations of inflation and output must then be very low to induce agents to choose this low expectations anyway, and, at some point, even *infinitely low* inflation and output gap will no longer decrease expectations enough to induce a deflationary spiral. Looking at Figure 9, this implies that as s^2 becomes smaller, the black line is considerably moved to the left, so that the deflationary spiral region decreases in size. When s^2 no longer satisfies (27) the deflationary spiral region disappears altogether and the target steady state is **globally stable** in the LTL model as well.

Finally, we return to the effect of the inflation target on the unstable steady state in the LTL model. As can be seen in Equation (25), the inflation level of the ZLB steady state depends linearly on the inflation target (π^T). This is illustrated in Figure 11. Here it can be seen that for a higher inflation target, the target steady state has an increased

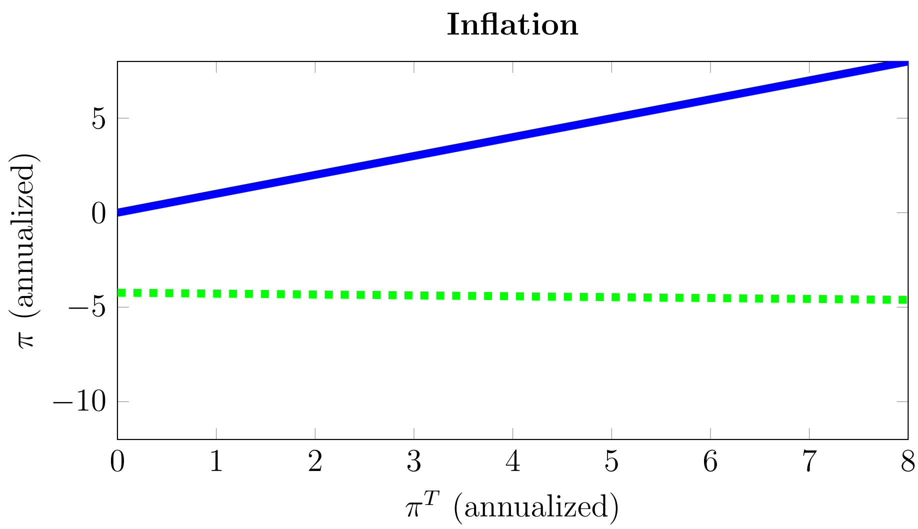

Figure 11: Bifurcation diagram of LTL model in π^T. The blue line represents the target steady state, and the green dashed line depicts the unstable ZLB saddle steady state. Points above the ZLB steady state imply convergence to the target steady state, while points below it lead to a deflationary spiral.

inflation value, while the ZLB steady state has a (slightly) decreased inflation value. An increased inflation target thus will make it less likely that negative shocks bring the model to the deflationary spiral region. This is also what we found in the model with 41 types in Figure 6. However, in the LTL model there is no lower bound on the expectation values agents consider, so (as long as s^2 satisfies (27)), the ZLB steady state cannot be made to disappear with a high enough inflation target.

6 Conclusion

We built a model with heterogeneous expectations to gain more insight in the possibility of self-reinforcing liquidity traps, and its relation with the anchoring of expectations. We found that when inflation and output gap expectations are relatively unanchored from their targets, bad shocks that drive the economy to the zero lower bound on the nominal interest can initiate a self-reinforcing wave of pessimism that keeps the economy in a liquidity trap. Whether the economy can recover from such a liquidity trap crucially depends on how unanchored expectations were when the bad shocks hit. Additionally, the central bank can reduce the likelihood of the occurrence of self-reinforcing liquidity traps by increasing its inflation target.

We conclude that it is crucial that in times where the zero lower bound is not binding, the central bank tries to let expectations become as strongly anchored to its targets as possible. In our model, this can be achieved by letting the interest rate respond aggressively enough to inflation and neither too weakly nor too strongly to output gap. The central bank then provides strong enough mean reversion to the target steady state to exclude a near unit root, so that almost self-fulfilling waves of optimism and pessimism cannot arise. As a consequence, inflation and output gap always remain relatively close to their target which, over time, anchors private sector expectations.

A 3 Type model

A.1 Proof Proposition 2

In a steady state where all agents are pessimists, (4) and (5) reduce to

$$x_t = (1 - \frac{\phi_2}{\sigma})(x^T - b) - \frac{\phi_1 - 1}{\sigma}(\pi^T - b - \pi^T) + \frac{\phi_2}{\sigma}x^T, \tag{A.1}$$

$$\pi_t = \beta E_t(\pi^T - b) + \kappa x_t. \tag{A.2}$$

Assume $\omega = \infty$. Then the steady state exists if and only if both output gap and inflation are such that in the next period again all agents are pessimists about both variables. This requires that both $x_t < x^T - \frac{b}{2}$ and $\pi_t < \pi^T - \frac{b}{2}$. Rewriting (A.1), the condition for output gap reduces to

$$x_t = x^T + b(\frac{\phi_2}{\sigma} - 1 + \frac{\phi_1 - 1}{\sigma}) < x^T - \frac{b}{2}, \tag{A.3}$$

which gives

$$\phi_1 < 1 + \frac{\sigma}{2} - \phi_2 \tag{A.4}$$

If this condition is satisfied (and thereby $x_t < x^T - \frac{b}{2}$), then it follows from (A.2) that

$$\pi_t < \beta E_t(\pi^T - b) + \kappa(x^T - \frac{b}{2}). \tag{A.5}$$

This implies that for $2\beta + \kappa > 1$, $\pi_t < \pi^T - \frac{b}{2}$ is automatically satisfied. Since κ is positive, and it reasonable to consider only calibrations with the discount factor satisfying $\beta > \frac{1}{2}$, we can conclude that when (A.2) holds, both $x_t < x^T - \frac{b}{2}$ and $\pi_t < \pi^T - \frac{b}{2}$, so that the pessimistic steady state exists. If (A.2) does not hold, then output gap expectations (and possibly also inflation expectations) will not stay pessimistic, so that the pessimistic steady state does not exist. Local stability of the steady state (and all other possible steady states under $\omega = \infty$) is proven in the online supplementary material.

A.2 Proof Proposition 3

In a steady state where all agents are pessimists, (11) and (12) reduce to

$$x_t = (x^T - b) + \frac{1}{\sigma}(\pi^T - b) + \frac{\bar{r}}{\sigma}, \tag{A.6}$$

$$\pi_t = \beta(\pi^T - b) + \kappa x_t. \tag{A.7}$$

Analogously to Appendix A.1 it follows that under these equations $x_t < x^T - \frac{b}{2}$ if and only if

$$\pi^T + \bar{r} < b(1 + \frac{\sigma}{2}), \tag{A.8}$$

and that $\pi_t < \pi^T - \frac{b}{2}$ then holds as well (assuming $2\beta + \kappa > 1$).

The steady state is however only consistent with equations (11) and (12) if it lies in the ZLB region of the model. This is the case if and only if

$$i_t = \pi^T + \bar{r} + \phi_1(\pi^T - b - \pi^T) + \phi_2(x^T - b - x^T) < 0, \tag{A.9}$$

or

$$\pi^T + \bar{r} < b(\phi_1 + \phi_2). \tag{A.10}$$

A.3 Proof Proposition 4

Inflation and output gap are again given by (A.7) and (A.6). Now it is required however, that $x_t < x^T - b$ and that $\pi_t < \pi^T - b$. The first inequality reduces to $\pi^T + \bar{r} < b$. If we assume that $\beta + \kappa > 1$, which will hold for most reasonable calibrations, $\pi_t < \pi^T - b$ then holds as well.

B Large type limit (LTL)

B.1 LTL and Bayesian updating

Equation (19) can be written as

$$E\pi_{t+1} - \pi^T = \frac{\int_{-\infty}^{\infty} b e^{-\omega(\pi_{t-1} - b - \pi^T)^2} e^{-\frac{b^2}{2s^2}} db}{\int_{-\infty}^{\infty} e^{-\omega(\pi_{t-1} - b - \pi^T)^2} e^{-\frac{b^2}{2s^2}} db} = \int_{-\infty}^{\infty} b\psi(b)db. \tag{B.1}$$

Here, $\psi(b) = \frac{e^{-\omega(\pi_{t-1} - b - \pi^T)^2} e^{-\frac{b^2}{2s^2}}}{\int_{-\infty}^{\infty} e^{-\omega(\pi_{t-1} - b - \pi^T)^2} e^{-\frac{b^2}{2s^2}} db}$ can be interpreted as agents' (posterior) probability density function of choosing bias b. This facilitates a comparison between the large type limit and Bayesian updating.

To make the relation between LTL and Bayesian updating more apparent, we multiply the numerator and denominator of $\psi(b)$ by $\frac{\sqrt{2\omega}}{\sqrt{2\pi}}$ and by $\frac{1}{s\sqrt{2\pi}}$ to get

$$\psi(b) = \frac{\frac{\sqrt{2\omega}}{\sqrt{2\pi}} e^{-\omega(\pi_{t-1}-b-\pi^T)^2} \frac{1}{s\sqrt{2\pi}} e^{-\frac{b^2}{2s^2}}}{\int_{-\infty}^{\infty} \frac{\sqrt{2\omega}}{\sqrt{2\pi}} e^{-\omega(\pi_{t-1}-b-\pi^T)^2} \frac{1}{s\sqrt{2\pi}} e^{-\frac{b^2}{2s^2}} db}. \tag{B.2}$$

We now obtain a perfect mapping to the posterior of a Bayesian updating process, given by

$$P(b|\pi_{t-1}) = \frac{P(\pi_{t-1}|b)P(b)}{\int_{-\infty}^{\infty} P(\pi_{t-1}|b)P(b)db} \tag{B.3}$$

First of all, the posterior $P(b|\pi_{t-1}) = \psi(b)$ gives the probability density of choosing bias b, given the observed value of π_{t-1}. The prior distribution of b is given by $P(b) = \frac{1}{s\sqrt{2\pi}} e^{-\frac{b^2}{2s^2}}$, which is the assumed distribution of the possible biases over the real line. The likelihood of observing realization π_{t-1}, given a value of the bias b, is now given by $P(\pi_{t-1}|b) = \frac{\sqrt{2\omega}}{\sqrt{2\pi}} e^{-\omega(\pi_{t-1}-b-\pi^T)^2}$. That is, agents believe that the data generating process (DGP) of π_{t-1} is equal to a constant, $\pi^T + b$, plus a normally distributed error term with variance $\frac{1}{2\omega}$. They then try to learn the true constant of the DGP by observing the noisy time series and applying Bayesian updating. Note that the true constant value in principle is allowed to change over time. What the agents try to infer in the above two equations is what the true value was likely to be at the moment that π_{t-1} was generated.

When ω goes to zero, the variance of the noise in the perceived DGP goes to infinity, so that the "likelihood" $\sqrt{\frac{\omega}{\pi}} e^{-\omega(\pi_{t-1}-b-\pi^T)^2}$ goes to zero for all values of π_{t-1}. This means that agents believe that no useful information about b can be inferred from observing π_{t-1}. As a consequence the posterior should just equal the prior, which is indeed the case according to Equation (B.2) (all types get equal weight which means that the distribution of mass over values on the real line is fully determined by the distribution of types over the real line).

When $\omega = +\infty$ the variance of the noise in the perceived DGP goes to zero and the likelihood $e^{-\omega(\pi_{t-1}-b-\pi^T)^2}$ becomes degenerate: it equals 1 when the observation (π_{t-1}) equals the true value $\pi^T + b$, and 0 for all other possible values of π_{t-1}. This implies that observing π_{t-1} is perceived as being fully informative about the true value of the constant. After observing π_{t-1} agents immediately know what the true constant value was, and the posterior distribution should be degenerate as well, with all its mass on the true value $\pi^T + b$. This is indeed what Equation (B.2) tells us, i.e., for infinite intensity of choice it holds that $\psi(b) = 1$ for $b = \pi_{t-1} - \pi^T$ and $\psi(b) = 0$ everywhere else (all agents switch with probability 1 to expecting the last observed value π_{t-1}).

We can conclude that if we interpret the LTL as Bayesian updating, then the intensity of choice parameter determines how informative observations of π_t are about the expectations that agents should choose. The larger the intensity of choice, the less "noise" agents think they encounter, and the more clear it is to them, what prediction value they should choose in the next period.

B.2 Proof Proposition 5

The Jacobian in the fundamental steady state equals

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

with

$$B = \frac{2\omega s^2}{2\omega s^2 + 1}. \tag{B.4}$$

The eigenvalues therefore are given by

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

Local stability requires $\lambda_1 < 1$ and $\lambda_2 > -1$. By keeping only the square root on one side of the equation and taking squares, $\lambda_1 < 1$ can be written as

$$\phi_1 > 1 + (-\beta B - \frac{1}{B} + 1 + \beta)\frac{\sigma}{\kappa} + (\beta B - 1)\frac{\phi_2}{\kappa}, \tag{B.5}$$

Filling in B from (B.4) results in Condition (23).

Similarly, $\lambda_2 > -1$ can be written as

$$\phi_1 < 1 + (1 + \beta + \beta B + \frac{1}{B})\frac{\sigma}{\kappa} - (1 + \beta B)\frac{\phi_2}{\kappa}, \tag{B.6}$$

from which Condition (24) can be obtained.

B.3 Proof Proposition 6

When the zero lower bound is binding the LTL model becomes

$$x_t = \frac{x^T}{2\omega s^2+1} + \frac{2\omega s^2}{2\omega s^2+1}x_{t-1} + \frac{1}{\sigma}\frac{\pi^T}{2\omega s^2+1} + \frac{1}{\sigma}\frac{2\omega s^2}{2\omega s^2+1}\pi_{t-1} + \frac{\bar{r}}{\sigma}, \tag{B.7}$$

$$\pi_t = \beta\frac{\pi^T}{2\omega s^2+1} + \beta\frac{2\omega s^2}{2\omega s^2+1}\pi_{t-1} + \kappa x_t. \tag{B.8}$$

Solving for the steady state of this model results in (25) and (26). Steady state output gap and inflation both are negative if and only if

$$\sigma(1+(1-\beta)2\omega s^2) - \kappa(2\omega s^2+1)2\omega s^2 < 0, \tag{B.9}$$

which can be rewritten as (27).

The eigenvalues of the system defined by (B.7) and (B.8) are given by

$$\lambda_{1,2} = \frac{\omega s^2}{2\omega s^2+1}\left((1+\beta+\frac{\kappa}{\sigma}) \pm \sqrt{\left(1+\beta+\frac{\kappa}{\sigma}\right)^2 - 4\beta}\right). \tag{B.10}$$

This implies that the steady state is an unstable saddle if and only if

$$\omega s^2 > \frac{1}{\beta - 1 + \frac{\kappa}{\sigma} + \sqrt{(1+\beta+\frac{\kappa}{\sigma})^2 - 4\beta}}, \tag{B.11}$$

which, after some algebraic manipulation, reduces to (27). Therefore, when (27) does not hold the system has a unique attractor that lies outside the ZLB region. This implies that from all initial conditions inflation and output gap will go towards this attractor and cross the zero lower bound. Recovery then always occurs.

When (27) holds, initial conditions below the stable eigenvector through the steady state given by (25) and (26) lead to ever decreasing inflation and output gap, while initial conditions above it lead to increasing inflation and output gap, and thereby to recovery. The slope of this eigenvector is given by (28).

References

Agliari, Anna, Domenico Massaro, Nicolò Pecora and Alessandro Spelta (2014), Inflation targeting, recursive inattentiveness and heterogeneous beliefs, Technical report.

Agliari, Anna, Nicolo Pecora and Alessandro Spelta (2015), 'Coexistence of equilibria in a new keynesian model with heterogeneous beliefs', *Chaos, Solitons & Fractals* **79**, 83–95.

Anderson, Simon P, Andre De Palma and Jacques François Thisse (1992), *Discrete choice theory of product differentiation*, MIT press.

Andrade, Philippe, Richard K Crump, Stefano Eusepi, Emanuel Moench et al. (2014), 'Noisy information and fundamental disagreement', *Staff Reports* **655**.

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

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.

Ball, Laurence M (2013), 'The case for four percent inflation', *Central Bank Review* **13**(2), 17.

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.

Branch, William A and Bruce McGough (2009), 'A new keynesian model with heterogeneous expectations', *Journal of Economic Dynamics and Control* **33**(5), 1036–1051.

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.

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

Clarida, Richard, Jordi Galı and Mark Gertler (1998), 'Monetary policy rules in practice: some international evidence', *european economic review* **42**(6), 1033–1067.

Curtin, Richard (2010), 'Inflation expectations and empirical tests', *inflation Expectations* **56**, 34.

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

De Grauwe, Paul and Yuemei Ji (2016), 'Inflation targets and the zero lower bound in a behavioral macroeconomic model'.

Di Bartolomeo, Giovanni, Marco Di Pietro and Bianca Giannini (2015), Optimal monetary policy in a new keynesian model with heterogeneous expectations!, Technical report, Department of Communication, University of Teramo.

Eusepi, Stefano (2007), 'Learnability and monetary policy: A global perspective', *Journal of Monetary Economics* **54**(4), 1115–1131.

Eusepi, Stefano (2010), 'Central bank communication and the liquidity trap', *Journal of Money, Credit and Banking* **42**(2-3), 373–397.

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, Seppo Honkapohja and Kaushik Mitra (2016), 'Expectations, stagnation and fiscal policy'.

Evans, George W, Seppo Honkapohja et al. (2009), 'Robust learning stability with operational monetary policy rules', *Monetary Policy Under Uncertainty and Learning* **13**, 145–170.

Gasteiger, Emanuel (2014), 'Heterogeneous expectations, optimal monetary policy, and the merit of policy inertia', *Journal of Money, Credit and Banking* **46**(7), 1535–1554.

Hommes, Cars H and Joep Lustenhouwer (2015), 'Inflation targeting and liquidity traps under endogenous credibility', *CeNDEF working paper* (15-11).

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.

Lustenhouwer, Joep (2017), Monetary and fiscal policy under bounded rationality and heterogeneous expectations, Technical report, Ph.D. dissertation at the University of Amsterdam.

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.

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

Pecora, Nicolò and Alessandro Spelta (2017), 'Managing monetary policy in a new keynesian model with many beliefs types', *Economics Letters* **150**, 53–58.

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

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

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