# Online Appendix of "Inflation Targeting and Liquidity Traps under Endogenous Credibility"

Cars Hommes*[a,b] and Joep Lustenhouwer[c]

[a]CeNDEF, University of Amsterdam

[b]Tinbergen Institute, Amsterdam

[c]Otto-Friedrich-Universität Bamberg

January 8, 2019

This document contains supplementary material meant for online publication only. In Section I microfoundations of our New Keynesian model with heterogeneous expectations are derived. Section II concerns results without a lower bound on the nominal interest rate, while proofs of propositions regarding the zero lower bound are presented in Section III. Finally, extra simulations for robustness are collected in Section IV

*Corresponding author: Cars Hommes, University of Amsterdam; Address: Roeteresstraat 11, 1018WB Amsterdam; tel:+31-20-5254256; Email: c.h.hommes@uva.nl

# I Microfoundations

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

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

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

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

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

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

where $\theta$ is the elasticity of substitution between the different goods.

The household $i$ then chooses consumption ($C_t^i$), labor ($H_t^i$), and holdings of real bonds ($b_t^i$), that households can trade between each other but that are in zero net supply, in order to maximize

$$\tilde{E}_t^i \sum_{s=t}^{\infty} \beta^{s-t} \left[\frac{(C_s^i)^{1-\sigma}}{1-\sigma} - \frac{(H_s^i)^{1+\eta}}{1+\eta} - \frac{\tau_b}{2}(b_s^i)^2\right], \tag{I.4}$$

subject to its budget constraint

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

where $\beta_t$ is the discount factor, $w_t$ the real wage rate, $i_t$ the nominal interest rate, $\pi_t = \frac{P_t}{P_{t-1}} - 1$ is the inflation rate, and $T_t^i$ real lump sum transfers to household

$i$, including profits from firms. $\tilde{E}_t^i$ represents the subjective expectation operator that differs over the households. As in Kurz et al. (2013), $\tau_b$ is a small cost that prevents excessive borrowing at the individual level. This replaces an institutional constraint, and excludes explosive borrowing from being an equilibrium solution.[1]

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

$$(C_t^i)^{-\sigma} = \lambda_t^i,$$
$$(H_t^i)^\eta = \lambda_t w_t,$$
$$\tau_b b_t^i + \lambda_t^i = \beta \tilde{E}_t^i \frac{\lambda_{t+1}^i (1+i_t)}{1+\pi_{t+1}},$$

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

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

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

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

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

where $\hat{C}_t = \frac{C_t - \bar{C}}{\bar{C}}$, with $\bar{C}$ the steady state value of consumption, and $\bar{r} = 1 - \frac{1}{\beta}$ is the steady state real (and due to a zero steady state inflation also nominal) interest rate.

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

[1] Note that households of different types (fundamentalists or naive) will make different consumption and labor decisions and hence have different demands for borrowing and saving, imply that some households borrow from other households. Moreover, different histories of having been naive or fundamentalist in past periods leads to many different current individual bond holdings among households. However, aggregate bond holdings in the economy are always equal to zero.

are derived by Massaro (2013).

Next, we deviate form Kurz et al. (2013), and use a property of the discrete choice model, $n_{h,t} = \frac{e^{bU_{h,t-1}}}{\sum_{h=1}^{H} e^{bU_{h,t-1}}}.$, which determines the fractions of agents in each period as in Brock and Hommes (1997). Under this model it is implicitly assumed that the probability to follow a particular heuristic next period is the same across agents, i.e., independent of the heuristic they followed in the past. This reflects the fact that our agents are not inherently different, but that each of them faces the same trade-off between becoming naive or fundamentalist each period. We assume agents know (have learned) that all agents have the same probability to follow a particular heuristic in the future, and that they know that consumption decisions only differ between households in so far as their expectations are different. In that case households' expectations about their own future consumption coincide with their expectations about the future consumption of any other agent, and therefore with aggregate consumption. That is, $\tilde{E}_t^i[\hat{C}_{t+1}^i] = \tilde{E}_t^i[\hat{C}_{t+1}]$, with $\hat{C}_{t+1} = \int_0^1 \hat{C}_{t+1}^i di$. Agents therefore realize they should base their current period consumption decision on expectations about future *aggregate* consumption. The Euler equation can then be written as

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

Market clearing in each good $j$ market imposes that

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

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

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

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

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

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

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

where we use the fact that bond holdings are in zero net supply.

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

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

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

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

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

where

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

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

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

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

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

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

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

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

This can be written as

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

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

Log linearizing gives

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

which can be written as

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

or recursively as

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

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

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

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

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

This can be log linearized to

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

from which it follows that

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

Plugging this into (I.24) gives

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

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

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

with

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

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

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

Inserting this in (I.29) results in

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

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

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

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

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

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

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

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

# II Monetary policy without the ZLB

## II.1 System dimension

The system defined by equations (5), (6) and (10) - (13) is six dimensional. First of all, next periods inflation and output gap ($\pi_{t+1}$ and $x_{t+1}$) are determined by the current values of these variables ($\pi_t$ and $x_t$), and by next periods fractions ($m_{t+1}^{\pi}$ and $m_{t+1}^{x}$). These four variables are however not enough to determine the future dynamics of the system since $m_{t+2}^{\pi}$ and $m_{t+2}^{x}$, which determine $\pi_{t+2}$ and $x_{t+2}$, depend on $\pi_{t-1}$ and $x_{t-1}$, and are therefore not determined by the above mentioned variables. It follows that the system must be six dimensional and that the state vector can be written as

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

or as

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

## II.2 Proof Proposition 1

Combining equations (10) and (11) with (5) and (6) gives

$$x_t = x^T + (1 - \frac{\phi_2}{\sigma})\frac{(1-m_t^x)}{2}(x_{t-1} - x^T) - \frac{\phi_1 - 1}{\sigma}\frac{(1-m_t^\pi)}{2}(\pi_{t-1} - \pi^T), \quad \text{(II.3)}$$

$$\pi_t = \beta\frac{(1+m_t^\pi)}{2}\pi^T + \beta\frac{(1-m_t^\pi)}{2}\pi_{t-1} + \kappa x_t. \quad \text{(II.4)}$$

In a steady state, (12) and (13) reduce to

$$m^x = tanh(-\frac{b}{2}(x - x^T)^2 \quad \text{(II.5)}$$

$$m^\pi = tanh(-\frac{b}{2}(\pi - \pi^T)^2) \quad \text{(II.6)}$$

This implies that when $x = x^T$ and $\pi = \pi^T$ fractions are given by $m^x = m^\pi = 0$. Next, we need to check that these steady state values are also consistent with (II.3) and (II.4). It can immediately seen that when $x_t = x_{t-1} = x^T$ and $\pi_{t-1} = \pi^T$, then (II.3) is satisfied. Plugging in $x_t = x$, $\pi_t = \pi_{t-1} = \pi^T = \pi$ and $m^x = m^\pi = 0$ in (II.4) we can rewrite the steady state Phillips curve as

$$x = \frac{1-\beta}{\kappa}\pi \quad \text{(II.7)}$$

This implies that as long as $x^T = \frac{1-\beta}{\kappa}\pi^T$, a steady state with $x = x^T$, $\pi = \pi^T$ and $m^x = m^\pi = 0$ exists.

## II.3 Jacobian and eigenvalues

In this section, first the Jacobian of the system given by (II.3), (II.4), (12) and (13) is presented. Next this Jacobian is evaluated at the fundamental steady state, and eigenvectors are derived.

The Jacobian is given by

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

with

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

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

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

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

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

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

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

with $x_t$ and $\pi_t$ given by (II.3) and (II.4)

In the fundamental steady state where $\pi_t = \pi^T$, $x_t = x^T$, and $m_t^x = m_t^\pi = 0$

for all t, the Jacobian reduces to

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

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

$$\lambda_1 = \frac{1}{4}\left( (1+\beta-\frac{\phi_2}{\sigma}-\kappa\frac{\phi_1-1}{\sigma}) + \sqrt{\left(1+\beta-\frac{\phi_2}{\sigma}-\kappa\frac{\phi_1-1}{\sigma}\right)^2 - 4\beta(1-\frac{\phi_2}{\sigma})}\right) \tag{II.8}$$

and

$$\lambda_2 = \frac{1}{4}\left( (1+\beta-\frac{\phi_2}{\sigma}-\kappa\frac{\phi_1-1}{\sigma}) - \sqrt{\left(1+\beta-\frac{\phi_2}{\sigma}-\kappa\frac{\phi_1-1}{\sigma}\right)^2 - 4\beta(1-\frac{\phi_2}{\sigma})}\right) \tag{II.9}$$

## II.4 Proof Proposition 3

We now analyze local stability of the steady state, and the bifurcations that arise, by considering first whether $\lambda_1 < 1$ and then whether $\lambda_2 > -1$. $\lambda_1 > -1$ and $\lambda_2 < 1$ are always satisfied.

### II.4.1 Conditions from $\lambda_1$

From (II.8) we know that $\lambda_1 > 1$ (so that the fundamental steady state is no longer locally stable) if and only if

$$\frac{1}{4}\left( (1+\beta-\frac{\phi_2}{\sigma}-\kappa\frac{\phi_1-1}{\sigma}) + \sqrt{\left(1+\beta-\frac{\phi_2}{\sigma}-\kappa\frac{\phi_1-1}{\sigma}\right)^2 - 4\beta(1-\frac{\phi_2}{\sigma})}\right) > 1 \tag{II.10}$$

This reduces to

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

or

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

which in terms of $\phi_1$ gives

$$\phi_1<\phi_1^{PF}=1-(2-\beta)\frac{\phi_2+\sigma}{2\kappa} \tag{II.11}$$

Below we show that a pitchfork bifurcation occurs at this value of $\phi_1$ by showing that two non-fundamental symmetric steady states are created here. In steady state (where $x^T=\frac{1-\beta}{\kappa}\pi^T$) (II.3) and (II.4) can be combined to

$$\left(1-\beta\frac{(1-m^\pi)}{2}+\kappa\frac{\frac{\phi_1-1}{\sigma}\frac{(1-m^\pi)}{2}}{(1-(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2})}\right)\pi=\left(1-\beta\frac{(1-m^\pi)}{2}+\kappa\frac{\frac{\phi_1-1}{\sigma}\frac{(1-m^\pi)}{2}}{(1-(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2})}\right)\pi^T \tag{II.12}$$

Non-fundamental steady states (where $\pi\neq\pi^T$) could exist as solutions of (II.12) if they satisfy

$$1-\beta\frac{(1-m^\pi)}{2}+\kappa\frac{\frac{\phi_1-1}{\sigma}\frac{(1-m^\pi)}{2}}{(1-(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2})}=0 \tag{II.13}$$

Writing this in terms of the inflation fraction gives

$$m^\pi=\frac{(2-\beta)(1-(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2})+\kappa\frac{\phi_1-1}{\sigma}}{-\beta(1-(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2})+\kappa\frac{\phi_1-1}{\sigma}} \tag{II.14}$$

The steady state values of $\pi$ then are

$$\pi^*=\pi^T\pm\sqrt{-\frac{2}{b}tanh^{-1}\left(\frac{(2-\beta)(1-(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2})+\kappa\frac{\phi_1-1}{\sigma}}{-\beta(1-(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2})+\kappa\frac{\phi_1-1}{\sigma}}\right)} \tag{II.15}$$

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

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

$$-1 \leq m^{\pi} < 0 \tag{II.16}$$

Since it is assumed that both $\sigma$ and $\phi_1$ are non-negative we must have

$$(1 - (1 - \frac{\phi_2}{\sigma})\frac{(1-m^x)}{2}) > 0 \tag{II.17}$$

Using this and (II.14), the inequalities in (II.16) reduce to

$$\beta(1 - (1 - \frac{\phi_2}{\sigma})\frac{(1-m^x)}{2}) - \kappa\frac{\phi_1 - 1}{\sigma} \geq (2-\beta)(1 - (1 - \frac{\phi_2}{\sigma})\frac{(1-m^x)}{2}) + \kappa\frac{\phi_1 - 1}{\sigma} > 0 \tag{II.18}$$

or equivalently

$$1 - \frac{\sigma}{\kappa}(1-\beta)\left(1 - (1 - \frac{\phi_2}{\sigma})\frac{(1-m^x)}{2}\right) \geq \phi_1 > 1 - \frac{\sigma}{\kappa}(2-\beta)\left(1 - (1 - \frac{\phi_2}{\sigma})\frac{(1-m^x)}{2}\right) \tag{II.19}$$

From the equivalence of (II.16) and (II.19), it can be concluded that as $\phi_1$ gets close to its right-hand limit, $m^{\pi}$ gets close to zero. This implies that $m^x$ , $x$ and $\pi$ also go to their fundamental values as this happens. Using that $m^x$ goes to zero in the limit, we see from Equation (II.19) that the limiting value of $\phi_1$ for which the non-fundamental steady states exist is

$$\phi_1^{PF} = 1 - \frac{\sigma}{\kappa}(2-\beta)(1 - (1 - \frac{\phi_2}{\sigma})\frac{1}{2}) = 1 - (2-\beta)\frac{\phi_2 + \sigma}{2\kappa} \tag{II.20}$$

At this point both steady states coincide with the fundamental steady state. We can conclude that at the bifurcation value indeed two non-fundamental steady states are created, which exists for values of $\phi_1$ larger than $\phi_1^{PF}$. The bifurcation therefore is a subcritical pitchfork bifurcation and the non-fundamental steady states must be unstable.

### II.4.2 Conditions from $\lambda_2$

Local stability disappears because $\lambda_2 < -1$ when

$$\frac{1}{4}\left((1 + \beta - \frac{\phi_2}{\sigma} - \kappa\frac{\phi_1 - 1}{\sigma}) - \sqrt{\left(1 + \beta - \frac{\phi_2}{\sigma} - \kappa\frac{\phi_1 - 1}{\sigma}\right)^2 - 4\beta(1 - \frac{\phi_2}{\sigma})}\right) < -1 \tag{II.21}$$

Rewriting this as

$$\sqrt{\left(1+\beta-\frac{\phi_2}{\sigma}-\kappa\frac{\phi_1-1}{\sigma}\right)^2-4\beta(1-\frac{\phi_2}{\sigma})} > 4+(1+\beta-\frac{\phi_2}{\sigma}-\kappa\frac{\phi_1-1}{\sigma}) \quad (II.22)$$

$$-4\beta(1-\frac{\phi_2}{\sigma}) > 16+8(1+\beta-\frac{\phi_2}{\sigma}-\kappa\frac{\phi_1-1}{\sigma}), \quad (II.23)$$

the condition reduces to

$$(2+\beta)\frac{\phi_2}{\sigma} > 3(2+\beta)-2\kappa\frac{\phi_1-1}{\sigma}, \quad (II.24)$$

which can be rewritten as

$$\phi_1 > 1+(2+\beta)\frac{3\sigma-\phi_2}{2\kappa}, \quad (II.25)$$

When one eigenvalue becomes $-1$, a 2-cycle must exists either below or above the bifurcation value. This makes the period doubling bifurcation either subcritical or supercritical. In what follows, $\phi_1$ is treated as the bifurcation parameter. The value of $\phi_2$ then turns out to determine if the bifurcation is subcritical or supercritical.

The 2-cycle in question is symmetric around the fundamental steady state. We thus have $(x_1-x^T)=-(x_2-x^T)$ and $(\pi_1-\pi^T)=-(\pi_2-\pi^T)$. Using this, (II.3) can be written as

$$x=x_1=x^T+(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2}(x_2-x^T)-\frac{\phi_1-1}{\sigma}\frac{(1-m^\pi)}{2}(\pi_2-\pi^T) \quad (II.26)$$

$$(1+(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2})x=-\frac{\phi_1-1}{\sigma}\frac{(1-m^\pi)}{2}(\pi_2-\pi^T)+(1+(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2})x^T \quad (II.27)$$

Plugging this in in (II.4), and using $x^T=\frac{1-\beta}{\kappa}\pi^T$, gives

$$\pi=\pi_1=\beta\frac{(1+m^\pi)}{2}\pi^T+\beta\frac{(1-m^\pi)}{2}\pi_2-\kappa\frac{\frac{\phi_1-1}{\sigma}\frac{(1-m^\pi)}{2}(\pi_2-\pi^T)}{(1+(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2})}+(1-\beta)\pi^T \quad (II.28)$$

$$\pi-\pi^T=-(\pi-\pi^T)(\beta\frac{(1-m^\pi)}{2}-\kappa\frac{\frac{\phi_1-1}{\sigma}\frac{(1-m^\pi)}{2}}{(1+(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2})}) \quad (II.29)$$

A 2 cycle (where $\pi \neq \pi^T$) must therefore satisfy

$$(1+\beta\frac{(1-m^\pi)}{2} - \kappa\frac{\frac{\phi_1-1}{\sigma}\frac{(1-m^\pi)}{2}}{(1+(1-\frac{\phi_2}{\sigma})\frac{(1-m^x)}{2})}) = 0 \tag{II.30}$$

In terms of the inflation fraction, this gives

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

In a 2-cycle around the fundamental steady state naive agents use the observation from period $t-1$ to give a prediction about period $t+1$. Therefore, in a 2-cycle they make no prediction errors, while fundamentalists do make prediction errors. In a 2-cycle we must therefore have

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

Now, if

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

the inequalities of (II.32) reduce to

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

If

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

(II.32) can never hold, and if

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

(II.32) reduces to

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

As $\phi_1$ comes close to making the right hand side of (II.34) or the left hand side of

(II.37) binding, the system comes close to the fundamental steady state. In the limit we therefore have $m^x = 0$. The limiting value of these restrictions therefore reduces to the bifurcation value

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

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

$$\phi_2 < 3\sigma, \qquad \text{(II.39)}$$

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

$$\phi_2 > 3\sigma, \qquad \text{(II.40)}$$

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

## II.5 Proof Proposition 3

The dynamical system given by (II.3) and (II.4) is linear in expectation fractions ($m_t^x$ and $m_t^\pi$). Furthermore, since the system with all fundamentalists is degenerate (a steady state is reached in every period), the Jacobian and eigenvalues for any given set of expectation fractions is scaled by the fraction of naive agents. It follows that if the linear system given by (II.3) and (II.4) is stable for all naive fractions, it is stable for any set of expectations fractions, which implies global stability of the fundamental steady state in our non-linear dynamical system. When $m_t^x = m_t^\pi = -1$ the system reduces to

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

$$\pi_t = \beta\pi_{t-1} + \kappa x_t \qquad \text{(II.42)}$$

The eigenvalues of this system are two times the eigenvalues given by (II.8) and (II.9). Now, replacing $\frac{1}{4}$ with $\frac{1}{2}$ in (II.10) and (II.21) and performing the same calculations as done in Appendix II.4.1 and II.4.2 gives the conditions given in the proposition.

# III Zero lower bound on the nominal interest rate

## III.1 Proof Proposition 4

From equations (17) and (18) it follows that in steady state, we must have

$$(1+m^x)x = \frac{(1+m^\pi)\pi^T + (1-m^\pi)\pi + 2\bar{r}}{\sigma} + (1+m^x)x^T, \tag{III.1}$$

and

$$\kappa x = \left(1 - \frac{1-m^\pi}{2}\beta\right)\pi - \frac{1+m^\pi}{2}\beta\pi^T \tag{III.2}$$

Combining these equations and using $x^T = \frac{1-\beta}{\kappa}\pi^T$ gives

$$\left((1+m^x)\left(1 - \frac{1-m^\pi}{2}\beta\right) - \frac{\kappa}{\sigma}(1-m^\pi)\right)\pi = \tag{III.3}$$
$$\left((1+m^x)\left(1-\beta+\frac{1+m^\pi}{2}\beta\right) + \frac{\kappa}{\sigma}(1+m^\pi)\right)\pi^T + \frac{\kappa}{\sigma}2\bar{r}$$

In a steady state in the ZLB region with infinite intensity of choice, differences in fractions can either be 0 or $-1$. Differences in fractions of 1 are not possible since in a steady state naive agents never make prediction errors. Moreover, $m^\pi = 0$ if and only if $\pi = \pi^T$. But in that case (III.3) reduces to $\pi^T = -\bar{r} < 0$, which contradicts the assumption of $\pi^T \geq 0$. Therefore $m^\pi = -1$ and $\pi \neq \pi^T$ must hold. Furthermore, if $\pi \neq \pi^T$ then it follows from the Phillips curve that $x \neq x^T$, so that it must be that $m^x = -1$ as well. The unique solution of (III.3) then is $\pi = -\bar{r}$. Corresponding output gap is given by $x = -\frac{1-\beta}{\kappa}\bar{r}$. This steady state only exists if it lies inside the ZLB region and hence implies a binding zero lower bound. It follows from (4) that this is the case if and only if

$$(\bar{r} + \pi^T)(1 - \phi_1 - (1-\beta)\frac{\phi_2}{\kappa}) < 0 \tag{III.4}$$

The first term in brackets is always positive. Therefore it follows that the steady state exists if and only if the Taylor principle is satisfied.

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

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

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

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

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

and are given by

$$\lambda_1 = \frac{1}{2}[1+\beta+\frac{\kappa}{\sigma} - \sqrt{(1+\beta+\frac{\kappa}{\sigma})^2 - 4\beta}] \tag{III.5}$$

$$\lambda_2 = \frac{1}{2}[1+\beta+\frac{\kappa}{\sigma} + \sqrt{(1+\beta+\frac{\kappa}{\sigma})^2 - 4\beta}] \tag{III.6}$$

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

## III.2 Proof Proposition 5

When expectations about both variables are fundamentalist, we have $E_t\pi_{t+1} = \pi^T$ and $E_t x_{t+1} = x^T$, and equation (4) implies that $i_t = \bar{r} + \pi^T > 0$. Therefore, the system is in the positive interest rate region by definition.

## III.3 Proof Proposition 6

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

## III.4 Proof Proposition 7

At least for the first two periods dynamics are given by the all naive expectations system of

$$x_{t+1} = x_t + \frac{\pi_t}{\sigma} + \frac{\bar{r}}{\sigma} \tag{III.7}$$

$$\pi_{t+1} = (\beta + \frac{\kappa}{\sigma})\pi_t + \kappa x_t + \frac{\kappa}{\sigma}\bar{r} \tag{III.8}$$

Iterating one more period gives

$$x_{t+2} = \left(1 + \frac{\kappa}{\sigma}\right) x_t + \left(\frac{1+\beta}{\sigma} + \frac{\kappa}{\sigma^2}\right)\pi_t + \left(\frac{2}{\sigma} + \frac{\kappa}{\sigma^2}\right)\bar{r} \tag{III.9}$$

$$\pi_{t+2} = \left(\left(\beta + \frac{\kappa}{\sigma}\right)^2 + \frac{\kappa}{\sigma}\right)\pi_t + \left((1+\beta)\kappa + \frac{\kappa^2}{\sigma}\right)x_t + \left((1+\beta)\frac{\kappa}{\sigma} + \frac{\kappa^2}{\sigma^2}\right)\bar{r} \tag{III.10}$$

In the all naive system, initial conditions above the stable eigenvector through the $\pi = -\bar{r}$, $x = -\frac{1-\beta}{\kappa}\bar{r}$ steady state lead output and inflation to move to the positive interest rate region, while lower initial conditions lead to divergence to minus infinity for both variables. Therefore, as long as expectations always remain naive for both variables, recovery occurs if and only if

$$x_t > -\frac{1-\beta}{\kappa}\bar{r} - \frac{1-(1-\beta)\frac{\sigma}{\kappa} + \sqrt{1+\frac{\sigma^2}{\kappa^2}\left((1+\beta)^2-4\beta\right)+2(1+\beta)\frac{\sigma}{\kappa}}}{2\sigma}\left(\pi_t+\bar{r}\right) \equiv \bar{x}^{ev} \tag{III.11}$$

We must however also consider the possibility that agents become fundamentalists in period $t+3$ about output gap or inflation. Naive expectations about output gap are excluded when

$$x_{t+2} = \left(1+\frac{\kappa}{\sigma}\right)x_t + \left(\frac{1+\beta}{\sigma}+\frac{\kappa}{\sigma^2}\right)\pi_t + \left(\frac{2}{\sigma}+\frac{\kappa}{\sigma^2}\right)\bar{r} > \frac{x_t+x^T}{2}$$

$$x_t > \frac{-\left(2(1+\beta)+2\frac{\kappa}{\sigma}\right)\pi_t - (4+2\frac{\kappa}{\sigma})\bar{r} + \sigma x^T}{2\kappa+\sigma} \equiv \bar{x}^{out} \tag{III.12}$$

When inflation expectations remain naive, a sufficient condition for recovery therefore is that (III.11) and (III.12) both hold.

If $\pi_t < \pi^T$, fundamentalist expectations about inflation only increase both variables for all subsequent periods, implying that above conditions that assume naive inflation expectations are still sufficient for recovery when inflation expectations could become fundamentalist. If $\pi_t > \pi^T$ inflation expectations are at least as large as $\pi^T$, so that sufficient conditions for recovery can be obtained by replacing $\pi_t$ with $\pi^T$ in (III.11) and (III.12). Therefore, a sufficient condition for recovery is

$$x_t > \max(\underline{x}^{ev}, \underline{x}^{out}), \tag{III.13}$$

with

$$\underline{x}^{ev} = -\frac{1-\beta}{\kappa}\bar{r} - \frac{1-(1-\beta)\frac{\sigma}{\kappa} + \sqrt{1+\frac{\sigma^2}{\kappa^2}\left((1+\beta)^2-4\beta\right)+2(1+\beta)\frac{\sigma}{\kappa}}}{2\sigma}\left(min(\pi_t,\pi^T)+\bar{r}\right) \tag{III.14}$$

$$\underline{x}^{out} = \frac{-\left(2(1+\beta)+2\frac{\kappa}{\sigma}\right)min(\pi_t,\pi^T) - (4+2\frac{\kappa}{\sigma})\bar{r} + \sigma x^T}{2\kappa+\sigma} \tag{III.15}$$

Next, we turn to sufficient conditions for divergence. From the above it follows that as long as inflation expectations remain naive, a sufficient condition for divergence is that (III.11) and (III.12) both do not hold. Inflation expectations remaining naive is guaranteed when $\pi_{t+2} < \frac{\pi_t + \pi^T}{2}$, which reduces to

$$x_t < \frac{-\left((\beta + \frac{\kappa}{\sigma})^2 + \frac{\kappa}{\sigma} - \frac{1}{2}\right)\pi_t + \frac{1}{2}\pi^T}{(1+\beta)\kappa + \frac{\kappa^2}{\sigma}} - \frac{r}{\sigma} \equiv \bar{x}^{inf} \tag{III.16}$$

A sufficient condition for a deflationary spiral therefore is

$$x_t < \min(\bar{x}^{ev}, \bar{x}^{out}, \bar{x}^{inf}), \tag{III.17}$$

## III.5 Proof Proposition 8

For $b = 0$, the system under the zero lower bound reduces to

$$x_{t+1} = \frac{x_t}{2} + \frac{x^T}{2} + \frac{\pi_t}{2\sigma} + \frac{\pi^T}{2\sigma} + \frac{\bar{r}}{\sigma} \tag{III.18}$$

$$\pi_{t+1} = (\beta + \frac{\kappa}{\sigma})(\frac{\pi_t}{2} + \frac{\pi^T}{2}) + \frac{\kappa}{2}x_t + \frac{\kappa}{2}x^T + \frac{\kappa}{\sigma}\bar{r} \tag{III.19}$$

The unique steady state of this system is

$$\pi = \frac{((2-\beta)\sigma + 2\kappa)\,\pi^T + 4\kappa\bar{r}}{(2-\beta)\sigma - 2\kappa} \tag{III.20}$$

$$x = x^T + \frac{2(2-\beta)}{(2-\beta)\sigma - 2\kappa}(\pi^T + \bar{r}) \tag{III.21}$$

It follows that $x$ and $\pi$ lie above the target if and only if the numerator in the fractions in (III.20) and (III.21) is positive. This is the case if and only if $\kappa < \frac{(2-\beta)\sigma}{2}$.

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

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2\sigma} \\ \frac{\kappa}{2} & \frac{\beta}{2} + \frac{\kappa}{2\sigma} \end{pmatrix} \tag{III.22}$$

Since the Jacobian is $\frac{1}{2}$ times the Jacobian of the system with all naive agents (given by (III.7) and (III.8)), it has the same eigenvectors. The eigenvalues are given by

$$\lambda_1 = \frac{1}{4}[1 + \beta + \frac{\kappa}{\sigma} - \sqrt{(1 + \beta + \frac{\kappa}{\sigma})^2 - 4\beta}] \tag{III.23}$$

$$\lambda_2 = \frac{1}{4}[1 + \beta + \frac{\kappa}{\sigma} + \sqrt{(1 + \beta + \frac{\kappa}{\sigma})^2 - 4\beta}] \tag{III.24}$$

Both eigenvalues lie in the unit circle if and only if $\kappa < \frac{(2-\beta)\sigma}{2}$, otherwise the steady state is a saddle point. In that case the slope of the stable eigenvector is given by (III.11).

# IV Robustness

## IV.1 Intensity of choice

Figure *1, below, plots the same simulation as in Figure 3 in the main body of the paper, but now for a higher intensity of choice parameter, $b$. In particular, the intensity of choice is set to 1 billion, which in this case is equivalent to an infinite intensity of choice, because the fractions of agents being fundamentalist jumps between 0 and 1, and not to intermediate values. It can be seen that in this case output and inflation dynamics are quite similar to those in Figure 3. That is, first a drift to positive inflation values arises, and then a downward drift that results in a deflationary spiral.

Figure *2 plots the case of a lower value of the intensity of choice. In particular the intensity of choice is set to 10000. As can be seen in the figure, in our economy, this intensity of choice is already low enough to make sure that credibility never comes above 0.6 or below 0.5, even when shocks bring inflation 3 per cent from target. Since the credibility does not fall significantly in such an inflationary episode, the large drifts in inflation that could be observed in Figure 3 do not arise. Instead inflation and especially output gap closely follow the thin green paths that would have arisen under rational expectations.

This analysis has shown that a higher intensity of choice, that implies stronger

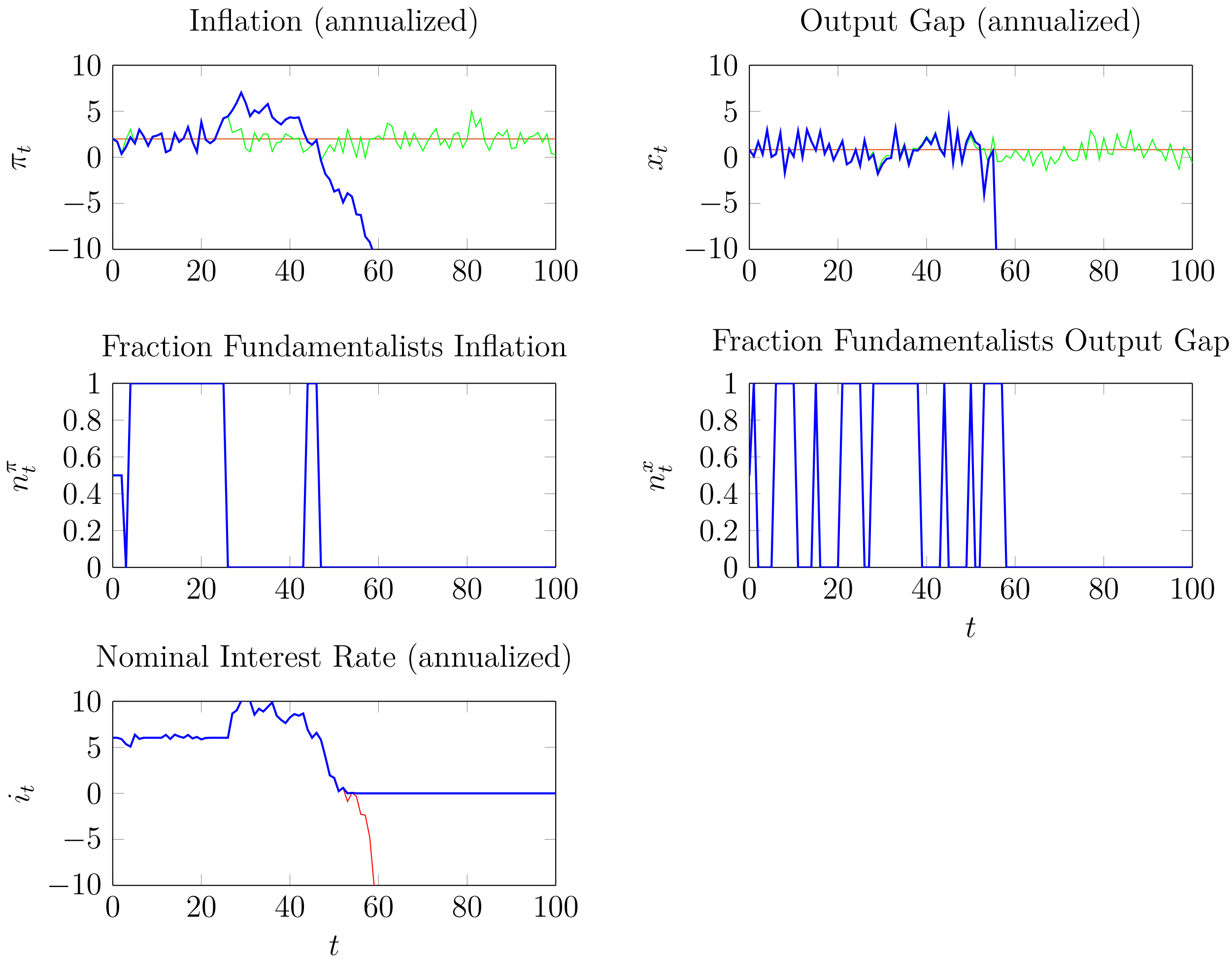


**Figure *1:** Simulated time series with intensity of choice of 1 billion. In the upper panels, the blue lines depict inflation and output in our model, the thin green lines depict time series that would have occurred under rational expectations, and the horizontal red lines depict the inflation and output gap targets. The bottom panel depicts both the actual interest rate (blue) and the rate prescribed by Equation (4) in the paper (thin red).

variations in credibility, does not lead to very large differences in dynamics, and that a lower intensity of choice leads to more stable dynamics, with deflationary spirals arising less easily, in line with the findings of Section 4. However, we consider our benchmark calibration, where there the credibility does not switch only between 0 and 1 but is also not stuck very close to 0.5 all the time, as the most realistic case.

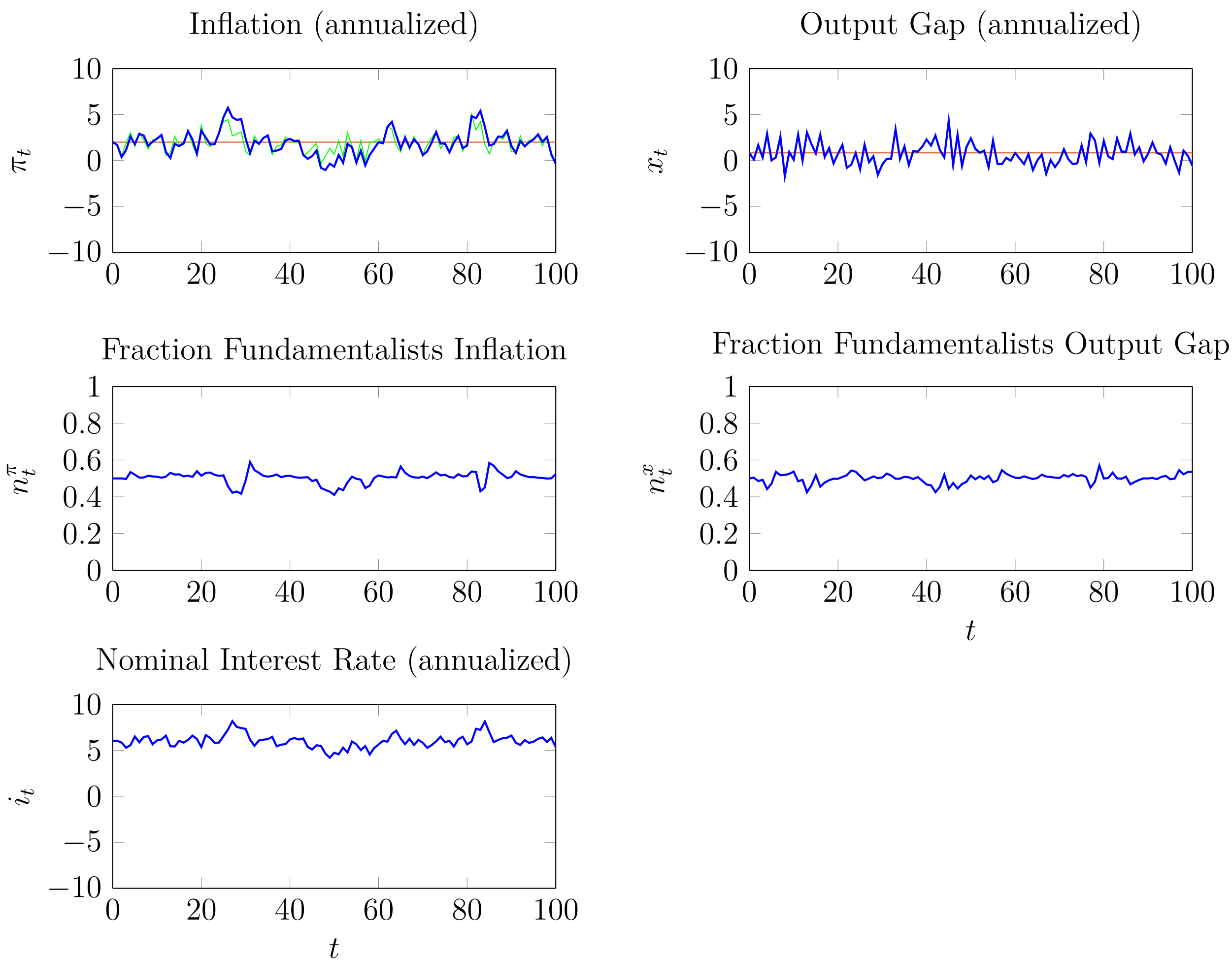


**Figure *2:** Simulated time series with intensity of choice of 10000. In the upper panels, the blue lines depict inflation and output in our model, the thin green lines depict time series that would have occurred under rational expectations, and the horizontal red lines depict the inflation and output gap targets.

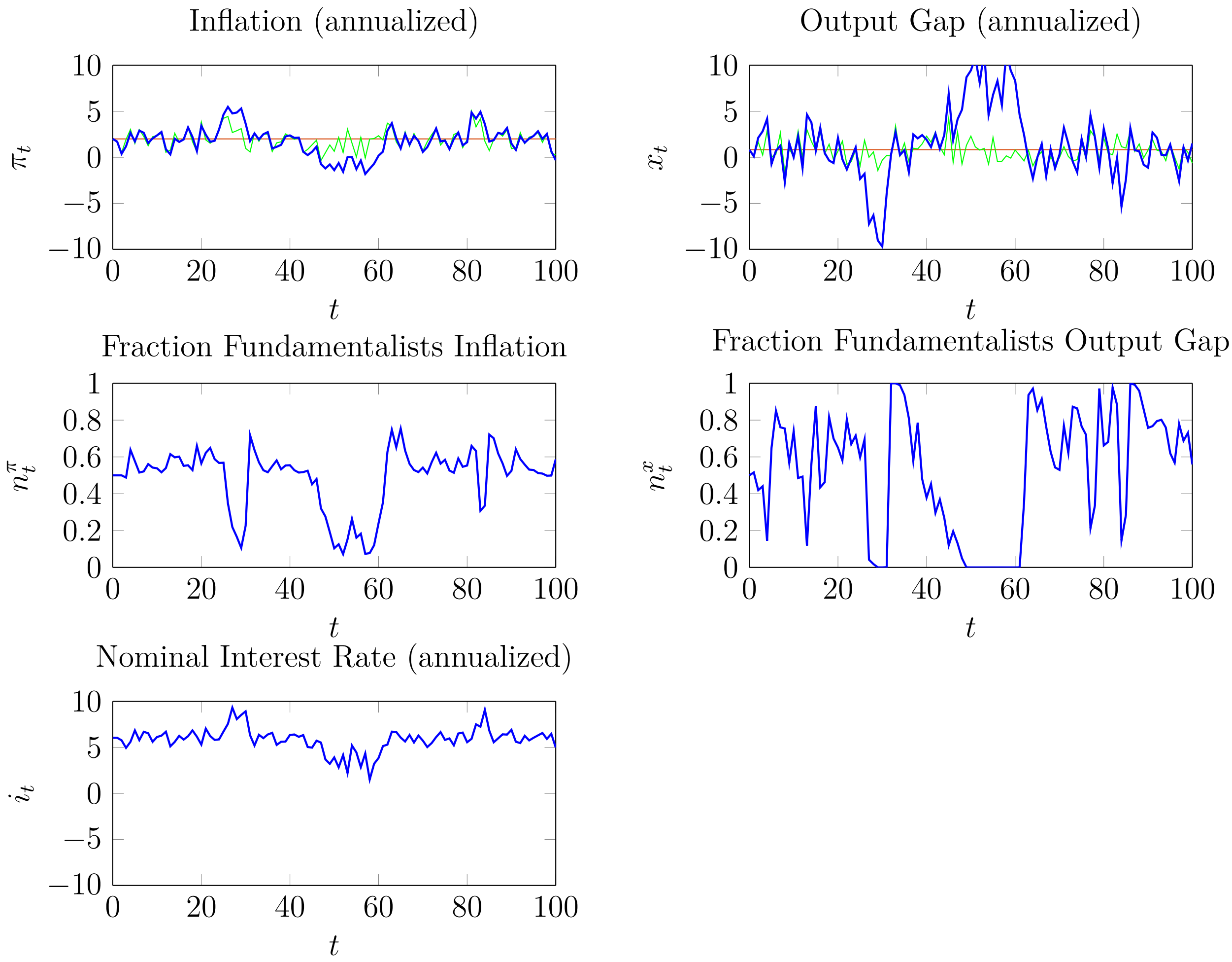


**Figure *3:** Simulated time series of model with aggressive inflation targeting ($\phi_1 = 1.5$). In the upper panels, the blue lines depict inflation and output in our model, the thin green lines depict time series that would have occurred under rational expectations, and the horizontal red lines depict the inflation and output gap targets. The bottom panel depicts both the actual interest rate (blue) and the rate prescribed by equation (4) (thin red).

## IV.2 Higher coefficient in Taylor rule

Since the danger of a liquidity trap and a deflationary spiral arises when the central bank loses credibility and when low self-fulfilling expectations allow inflation to drift downward, instead of the above measures, the central bank could also try to prevent these drifts altogether and try to always maintain a measure of credibility. It could do so by responding more aggressively to any deviation of inflation form target. That is, by increasing its coefficient on inflation in the interest rate rule.

Figure *3 plots the case where $\phi_1 = 1.5$. This can either be interpreted as reacting more strongly to inflation than would have been optimal without the zero lower bound, or as policy with a weight on output gap in the loss function of approximately $\mu = 0.007$. This reflects that in light of the liquidity trap analysis above, it is much more important to stabilize inflation than it is to stabilize output gap. This point was also made in the analytical analysis in Section 4, where we concluded that inflation expectations play a much large role than output gap expectations in determining whether or not the economy can recover from a liquidity trap.

As a result of the higher inflation coefficient, the interest rate in Figure *3 tracks short term inflation fluctuations much more closely than in the other figures. Any negative shock to inflation is immediately countered by a very low interest rate in the next period, which increases inflation again. As a result, drifts in inflation are less severe and credibility is never fully lost. While inflation still goes down somewhat between period 50 and 60, the zero lower bound never becomes binding

Note that the cost of the increased inflation coefficient in the Taylor rule arise in the form of stronger output gap fluctuations (also in times where no liquidity trap is imminent) than in Figure 3 in the main body of the paper, which is consistent with the fact that we are now considering policy derived from a loss function with a very low weight on output gap. These large output gap fluctuations imply a welfare loss.

# References


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

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

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

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

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