Experience-Based Heterogeneity in Expectations and Monetary Policy*

Lucas Radke† Florian Wicknig‡

September, 2021

For the latest version click here.

Abstract

We show within a New Keynesian model with experience-based learning (EBL) that heterogeneous expectations across age groups impair the ability of monetary policy to stabilise the economy. While experience effects on expectations reduce the transmission of monetary policy on inflation, they attenuate the stabilisation trade-off that monetary policy faces when aiming to stabilise both inflation and the output-gap under supply shocks. Moreover, under EBL, a variation in the age distribution has a composition effect on aggregate expectations which are a size-weighted average over cohort-specific expectations. As the share of old individuals increases, the composition effect increases the transmission of monetary policy on inflation but aggravates its stabilisation trade-off under supply shocks.

Keywords: Monetary Policy, Learning, Heterogeneous Expectations, Experience Effects, Demography

JEL Classification: D84, E31, E52, E70

*The authors thank their supervisors Johannes Pfeifer and Andreas Schabert. Moreover, we thank Harris Dellas, Wouter den Haan, Emanuel Hansen, Pasquale Foresti, Matteo Iacoviello, Ricardo Reis, Paul Schempp, David Finck, Matthias Pelster, Jenny Chan, Emanuel Gasteiger and Thomas Winberry as well as participants at the 2nd Behavioural Macroeconomics Workshop (Bamberg), the European Economic Association Conference (2020), the 2nd Conference on Behavioral Research in Finance, Governance, and Accounting (2020), the Spanish Economic Association's 2020 Meeting, the Royal Economic Society (2021) and the Society for Computational Economics (2021) for helpful comments.

†Radke: Center for Macroeconomic Research, University of Cologne, Albertus-Magnus-Platz, 50923 Cologne, Germany; Email: radke@wiso.uni-koeln.de.

‡Wicknig (Corresponding Author): Center for Macroeconomic Research, University of Cologne, Albertus-Magnus-Platz, 50923 Cologne, Germany; Email: fwicknig@wiso.uni-koeln.de.

1. Introduction

Private sector expectations, which are a key determinant for the implementation of monetary policy, exhibit substantial cross-sectional heterogeneity (e.g. Branch, 2004; Pfajfar and Santoro, 2010). But despite heterogeneity in expectations altering the propagation of shocks and transmission of monetary policy (e.g. Branch and McGough, 2011, 2018; De Grauwe, 2011; Massaro, 2013), the standard New Keynesian model still employs rational and thus homogeneous expectations.

In the present paper, we relax this assumption and endogenise expectational heterogeneity in an overlapping generations New Keynesian model featuring *Experience-Based Learning* (EBL) (Malmendier and Nagel, 2016).[1] Expectations are a function of the different economic experiences that individuals made over their lifetime. The resulting expectational heterogeneity across age groups can considerably impair the ability of monetary policy to stabilise the economy. Moreover, since aggregate expectations are a size-weighted average over cohort-specific expectations, a variation in the demographic structure affects aggregate expectations through a composition effect that we call the *Experience Channel*.

We show that, under EBL, the pass-through of monetary policy on expectations weakens relative to models without experience-effects on expectations. As a result, monetary policy overstates its influence on inflation if experience effects on expectations are neglected. At the same time, the stabilisation trade-off that monetary policy faces under supply shocks attenuates under EBL. Further, demographic shifts affect monetary policy through the Experience Channel that considerably increases the transmission of monetary policy on inflation in older societies. As a result, the trade-off between inflation and output gap stabilisation aggravates. Hence, when monetary policy wants to assess how a variation in the demographic structure affects its transmission on inflation, it should take into account experience effects on expectations.

Experience-Based Learning. Malmendier and Nagel (2016) provide empirical evidence that differences in inflation expectations across age groups are largely driven by differences in their *experienced* inflation history.[2] Figure 1 illustrates that young cohorts' expectations are more sensitive to recent observations than those of old individuals.[3] The markers denote one-year ahead inflation expectations of a "young" (red) and an "old"

[1] This contrasts with earlier studies that introduced expectational heterogeneity either by exogenously dividing the population into individuals with different forecasting models or endowing agents with a discrete choice from a finite set of predictors.

[2] Experiences are also relevant for individuals' risk taking behaviour or investment decisions (see e.g. Kaustia and Knüpfer, 2008; Malmendier and Nagel, 2011). Malmendier et al. (2020a) show that inflation experiences of Federal Open Market Committee members affect their inflation expectations.

[3] We extend Figure I from Malmendier and Nagel (2016) by plotting the annual percentage change of the seasonally-adjusted U.S.-CPI to facilitate the comparison of *expected* inflation to *experienced* inflation.

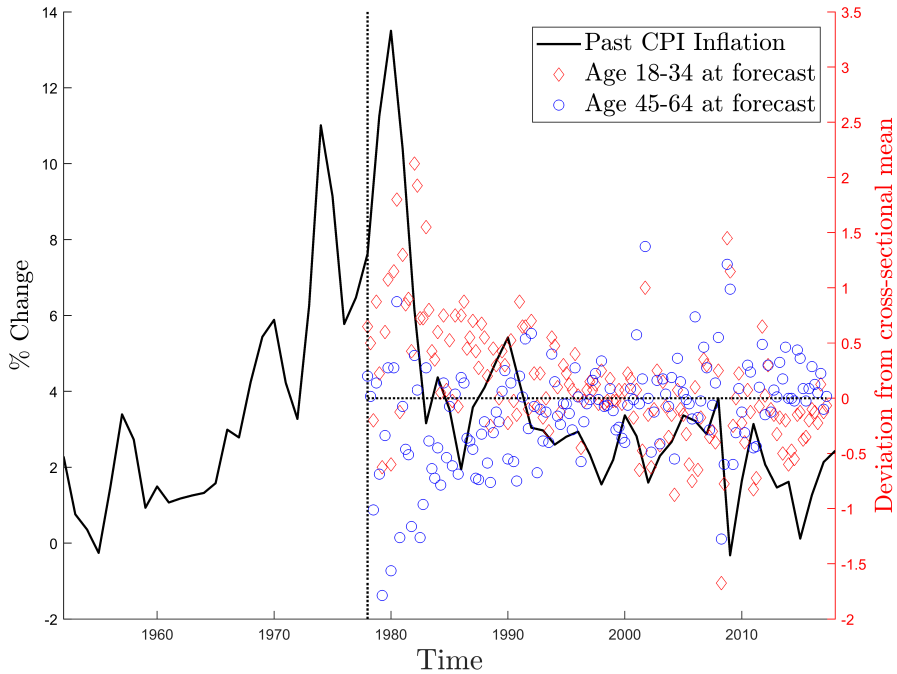

FIGURE 1: Realised Inflation vs. One-Year Ahead Inflation Expectations Across Cohorts

Notes: left y-axis: annual percentage change of the seasonally-adjusted U.S.-CPI (solid black line). Right y-axis: quarterly 4-quarter moving average of cohort inflation expectations, expressed as percentage deviations from from the cross-sectional mean (markers). We take a moving-average to concentrate on lower frequency variation. Inflation expectations are based on the Michigan Survey of Consumers (question: *Expected Change in Prices During the Next Year*) from 1978Q1 to 2017Q4. Cohorts define persons of a certain age group at a specific point in time, i.e. no age group is tracked over time.

(blue) cohort as deviation from the cross-sectional mean (percentage points, right y-axis). The black solid line depicts the year-on-year realised CPI-inflation from the previous year (left y-axis). The heterogeneity in expectations across different age groups is particularly pronounced in the early 1980s, because young individuals whose whole inflation history consists of the high inflation rates during the 1970s tend to have higher inflation expectations than those individuals who also observed low inflation during the 1950s and 1960s. Importantly, the high inflation expectations of young agents are not a function of their age but of their lifetime experience as can be inferred from the reversal of the ordering at the end of the sample where recent inflation experiences were extremely low.

A New Keynesian Model with EBL. The present paper embeds EBL into a New Keynesian model with overlapping generations à la Blanchard (1985) and Yaari (1965). We assume that agents behave like econometricians who form expectations about future economic variables based on forecasting models whose parameters they constantly revise as new data becomes available. In particular, agents forecast future variables with a covariance stationary auto-regressive process of order one with time-varying AR parameter,

which we denote as an agent's *perceived persistence*. Following the empirical analysis of Malmendier and Nagel (2016), agents put more weight on recently observed data points rather than those observed early in life, while ignoring any data *prior* to their birth. The weight attached to new observations when updating beliefs decreases in age, rendering young individuals more sensitive to new data points than older ones. It is the specification of the weight by which EBL differs from standard approaches in the learning literature like Constant-Gain-Learning (CGL) where all cohorts attach the same constant weight to new information so that expectations are homogenous across cohorts.

Our calibrated model with EBL generates a quantitatively substantial heterogeneity in expectations across age groups. The expectational heterogeneity in our model stems from differences in the perceived persistence different cohorts attach to economic variables. On average, the perceived persistence is more dispersed for young cohorts, because the parameter estimates of their forecasting rules are based on fewer observations, while more recent observations are overweighted. As young agents rely on few data points, recurrent reversals in inflation and the output gap make them perceive both variables to be less persistent, on average.

Transmission of Monetary Policy. We show that EBL endogenously reduces the aggregate perceived persistence in the economy relative to CGL, because individuals only use life-time information to forecast macroeconomic variables. The decrease in the aggregate perceived persistence reduces the impact of monetary policy on expectations. Intuitively, under adaptive expectations, monetary policy has a delayed impact on expectations by influencing current macroeconomic variables which are used to revise beliefs only tomorrow. The aggregate perceived persistence can then be interpreted as the weight agents attach to *past* monetary policy actions. As a consequence of the lower impact of past monetary policy actions on current variables, the transmission of monetary policy on inflation is impaired under EBL.[4] Hence, by neglecting experience effects on expectations, monetary policy *overstates* its impact on inflation. Moreover, a model with CGL cannot replicate the endogenous reduction in the aggregate perceived persistence under EBL unless the weight individuals attach to new information is set to an empirically implausibly high value.

We further show that under supply shocks, a higher weight on stabilising the output-gap increases inflation volatility by less if experience effects on expectations are considered. Hence, under EBL, monetary policy faces a weaker stabilisation trade-off than under CGL. However, the lower trade-off follows from the impaired effectiveness of monetary policy to affect expectations under EBL.

[4] As shown in Slobodyan and Wouters (2012a), the aggregate perceived persistence in the economy affects the response of the endogenous variables in response to exogenous disturbances.

Variations in the Demographic Structure. In response to an increase in the share of old individuals, the aggregate perceived persistence considerably increases under EBL, while it is hardly affected under CGL. The different response of the aggregate perceived persistence can be ascribed to the Experience Channel. Since under EBL the perceived persistence is heterogeneous across age groups, a variation of the age distribution directly affects the aggregate perceived persistence through a composition effect. Due to the Experience Channel, a higher share of old individuals increases the weight individuals attach to past monetary policy actions when forming expectations. Thereby, the transmission of monetary policy on inflation increases while its stabilisation trade-off under supply shocks aggravates.[5] Hence, in old societies, monetary policy gets more influence on inflation but any reduction in the volatility of inflation is associated with a stronger increase in output gap volatility. If monetary policy neglects the Experience Channel, it understates the effect of a variation in the age distribution on its transmission to inflation.

Related Literature. Our work is related to several strands of literature.
First, we relate to the adaptive learning literature that studies monetary policy within a New Keynesian model as surveyed in Eusepi and Preston (2018). In contrast to this literature, we assume that individuals are agnostic about the minimum state variable representation of the economy. Instead, they form expectations based on simple autoregressive forecasting models as indicated by the empirical evidence of Malmendier and Nagel (2016). We further assume that the parameters of the forecasting model are recursively updated and differ across age groups so that expectations are heterogeneous across generations.
Second, we contribute to the literature that analyses monetary policy within a New Keynesian model with heterogeneous expectations. This literature is mostly concerned with determinacy properties in the presence of heterogeneous expectations (see e.g. Branch and McGough, 2009; Gasteiger, 2014; Massaro, 2013). Expectation-formation in these studies is time-invariant, whereas our approach allows for real-time updates. Further, these studies take expectation heterogeneity as given, whereas in our model it arises endogenously from cohorts that had different lifetime experiences. This assigns a role for the demographic structure to affect the conduct of monetary policy.
Third, there is a literature analysing experience-based expectation heterogeneity in theoretical models. However, most focus on asset pricing in partial equilibrium, as Collin-Dufresne et al. (2016), Ehling et al. (2018), Malmendier et al. (2020b), Nagel and Xu

[5] This seems to stand in contrast with the Japanese experience of an older society but limited power of conventional monetary policy. However, the channels through which demography matters for this result, e.g. a saving glut induced by workers saving for retirement, is different from ours (see Carvalho et al., 2016). We instead introduce a novel channel of how demography and monetary policy pass-through are linked.

(2019), and Schraeder (2015). The only model using EBL in a general equilibrium framework that we are aware of is Acedański (2017), who explores its implication on the wealth distribution. The present paper is the first one to investigate the impact of the Experience Channel on monetary policy.

Lastly, we contribute to the growing literature that studies the link between demographic changes and monetary policy. There is broad agreement that, in the long-run, longevity and declining birth rates contribute to a reduction in the real interest rate (e.g. Aksoy et al., 2019; Eggertsson et al., 2019; Kara and von Thadden, 2016), which carries the risk of a binding zero-lower bound. Wong (2019) and Berg et al. (2020) consider how the transmission of monetary policy differs across age groups. Both studies deliver different conclusions on the efficacy of monetary policy in an ageing society. Baksa and Munkácsi (2020) explore the relation between ageing, inflation and optimal monetary policy. Our work points to a hitherto unconsidered channel of how shifts in demography affect the transmission of monetary policy through experience effects.

Structure. Section 2 presents our NK-model with an overlapping generations structure. In section 3 we explain the details of EBL. Our results concerning the implications of EBL and monetary policy are discussed in section 4. Finally, section 5 concludes.

2. The Model

In the present section, we consider a New Keynesian framework. However, we deviate from the standard model (e.g. Galí, 2015) by assuming that households form expectations based on their lifetime experiences. Conceptually, we follow the statistical learning (SL, henceforth) literature that assumes agents to forecast future outcomes by using a simplified model of the economy not knowing its actual law of motion. Agents constantly revise their beliefs as new observations become available. Since we allow this revision of beliefs to depend on an agent's age, we assume households to face a constant probability of death á la Blanchard (1985) and Yaari (1965).[6] Their saving decision involves the formation of inflation and output gap expectations, which depend on their age and lifetime experiences. This is the only source of heterogeneity across agents. Taken by itself, heterogeneous expectations lead to different saving decisions across agents. However, similar to other literature using boundedly rational agents (e.g. Mankiw and Reis, 2007, Adam et al., 2016 and Ehling et al., 2018) we abstract from the wealth distribution as an additional state to focus on the effects stemming from EBL via the expectation operator. To achieve this we employ the approach of Branch and McGough (2009) as discussed below.

[6] A similar assumption on demography is made by Ehling et al. (2018) or Galí (2021). Note further that we exclude the possibility of individuals to retire after their working life to keep the model simple.

Households own intermediate good producers that use labour to provide an input to a competitive final good producer whose output households consume. Monetary policy influences the bond rate according to a Taylor (1993)-rule.

2.1. Households

Households provide labour, form expectations according to EBL and own intermediate good firms. At each point in time, the mass of households is constant and normalised to one. They face an age-independent probability, $\omega \in [0,1]$, of surviving into the following period. In turn, at the beginning of each period a share of $1-\omega$ households deceases and is replaced by new-born households of equal mass. Consequently, the mass of a cohort born in period k at time $t \geq k$ is given by $(1-\omega)\omega^{t-k}$.
A household born in period k maximises the discounted sum of lifetime utility:

$$\tilde{E}_t^k \sum_{j=0}^{\infty} (\beta\omega)^j \, u(c_{t+j|k}, l_{t+j|k}) \ , \tag{1}$$

subject to the sequence of period budget constraints:

$$p_t c_{t|k} + b_{t|k} = r_{t-1}(b_{t-1|k} + z_{t|k}) + p_t w_t l_{t|k} + \mathcal{D}_{t|k} \ , \tag{2}$$

where β denotes the discount factor, p_t the price level and $\tilde{E}_t^k$ denotes the subjective expectations operator, which potentially differs across cohorts k and is specified below. Households from cohort k receive labour income which is the product of nominal hourly wages, $p_t w_t$, and working hours in cohort k, $l_{t|k}$. They invest in private one-period nominal bonds, $b_{t|k}$, which pay nominal interest rate r_t, known as of t, tomorrow. We assume each household owns equal shares in every firm so that nominal dividends are equal across cohorts, i.e. $\mathcal{D}_{t|k} = \mathcal{D}_t$.
The time of death is uncertain and households may die with wealth. To avoid the inefficiency of accidental bequests we follow Blanchard (1985) and introduce insurance companies that pay bond shares as annuity payments $z_{t|k}$ and that receive all assets at the time of death. Profits for a particular company contracting with cohort k are:

$$\pi_t^I = (1-\omega)\, b_{t-1|k} - \omega z_{t|k} \ .$$

Due to free entry insurers make zero-profits so that $z_{t|k} = \frac{1-\omega}{\omega} b_{t-1|k}$. The above sequence of period budget constraints is supplemented with a solvency condition of the form:

$$\lim_{T\to\infty} \tilde{E}_t^k \left\{ \mathcal{R}_{t,T} b_{T|k} \right\} = 0 \ , \tag{3}$$

where $\mathcal{R}_{t,T} = (\prod_{s=t+1}^{T} r_s)^{-1}$. We assume the following form of the felicity function:

$$u(c_{t|k}, l_{t|k}) = \ln\left(c_{t|k}\right) + \psi_n \ln\left(1 - l_{t|k}\right) \ ,$$

where ψ_n is a utility weight. Maximising (1) subject to (2) yields the optimal consumption/saving decision:[7]

$$1 = \tilde{E}_t^k \left\{ \beta \frac{p_t c_{t|k}}{p_{t+1} c_{t+1|k}} r_t \right\} \ . \tag{4}$$

Equation (4) denotes the household's Euler equation. While households of all ages face the same nominal interest rate, they have different expectations of the real rate. Hence, a household expecting a high future return, saves more and values future consumption more than a household whose past experiences make her believe in dismal real future returns.[8] Notwithstanding that age-related heterogeneity in expectations implies differences in cohort wealth, we aggregate the economy without considering the wealth distribution as an additional state variable as outlined in Section 2.3.

Furthermore, we follow Evans and Honkapohja (2012) and Slobodyan and Wouters (2012a) that require near-rational agents to forecast variables only one period ahead (e.g. of variables in their Euler equation so that the approach is called *Euler equation learning*). Hence, the Euler equation for the current period presents households' decisions as a function of the expected state of the economy tomorrow only.[9]

We also derive the labour supply of a household from cohort k that, via different consumption choices among cohorts, is cohort specific:

$$\psi_n \frac{c_{t|k}}{(1 - l_{t|k})} = w_t \ . \tag{5}$$

2.2. Firms

There are two types of firms. Final good firms use intermediate inputs to provide an aggregate consumption good. Intermediate good firms are owned by households and operate on a monopolistically competitive market. The choice to set up firms as in the usual NK-model serves to make our departure from the standard case minimal.

[7] Derivations are delegated to the Internet Appendix.

[8] Malmendier and Nagel (2011) provide empirical evidence that experience effects in individuals' expectation formation contribute to differences in their savings decision.

[9] For further details on the implication of Euler-equation learning on agents behaviour see Appendix B and Evans et al. (2013). Other approaches assume agents form forecasts on longer horizons, e.g. Preston (2005).

2.2.1. Final Good Firm

The aggregate consumption good in the economy, y_t, is produced by a perfectly competitive firm which is aggregating intermediate goods $i \in [0,1]$ produced by intermediate firms according to the technology:

$$y_t = \left[\int_0^1 y_{i,t}^{\frac{\varepsilon-1}{\varepsilon}} di\right]^{\frac{\varepsilon}{\varepsilon-1}} , \tag{6}$$

where $\varepsilon > 0$ is the elasticity of substitution among the intermediate goods, $y_{i,t}$. The final good firm chooses the quantities of intermediate goods to maximise its profits. The demand for intermediate good i is given by:

$$y_{i,t} = \left(\frac{p_{i,t}}{p_t}\right)^{-\varepsilon} y_t , \tag{7}$$

where $p_{i,t}$ denotes the price at which the intermediate good firm i sells the input to final good producers.

2.2.2. Intermediate Good Firms

Each household alive in period t owns an equal share in each intermediate good firm $i \in [0,1]$ that produces a differentiated good on a monopolistically competitive market. Since all households are involved in firms to an equal degree, the latter have average expectations as detailed below. We assume that the share of a deceasing household is transmitted to a new-born one instantaneously. Production of intermediate good i follows the technology:

$$y_{i,t} = l_{i,t}^{\alpha} , \tag{8}$$

where $l_{i,t}$ denotes labour demand of firm i and $0 < \alpha \leq 1$. Intermediate firm i sells its good at price $p_{i,t}$ but, when changing its price, pays quadratic nominal price adjustment costs à la Rotemberg (1982). Hence, the firm faces an inter-temporal problem that stems from the effect of $p_{i,t}$ on future price adjustment costs. The costs of changing prices are proportional to the nominal value of aggregate production:

$$\frac{\phi}{2}\left(\frac{p_{i,t}}{p_{i,t-1}} - 1\right)^2 p_t y_t,$$

where ϕ measures the degree of nominal rigidity. The adjustment cost increase in the scale of price changes and in the size of economic activity. Current real period profits,

$d_{i,t} = \frac{\mathcal{D}_{i,t}}{p_t}$, of firm i are given by:

$$d_{i,t} = \frac{p_{i,t}}{p_t} y_{i,t} - w_t l_{i,t} - \frac{\phi}{2} \left(\frac{p_{i,t}}{p_{i,t-1}} - 1 \right)^2 y_t \ .$$

Taking aggregate prices as given, firm i chooses $p_{i,t}$ and $y_{i,t}$ to maximise discounted profits:

$$\max_{p_{i,t+j}, y_{i,t+j}} \bar{E}_t \sum_{j=0}^{\infty} \omega^j Q_{t,t+j} \ d_{i,t+j} \ ,$$

subject to the demand schedule of final good firms (7). The expectation operator $\bar{E}_t \mathbf{z}_{t+1} \equiv (1-\omega) \sum_{k=-\infty}^{t} \omega^{t-k} \tilde{E}_t^k \mathbf{z}_{t+1}$ denotes the aggregated expectations across all cohorts alive in period t for a generic variable $\mathbf{z}$ and is a size-weighted sum of cohort expectations. Note that the generational structure matters for aggregating the decisions of the households of different age and especially when aggregating the expectations of differently aged households. Since households hold equal shares in every firm, firms use a weighted average of household expectations.[10] Further, $Q_{t,t+j} \equiv \beta^j \frac{c_t}{c_{t+j}}$ denotes the aggregate real stochastic discount factor of households, where $c_t = (1-\omega) \sum_{k=-\infty}^{t} \omega^{t-k} c_{t|k}$.

2.2.3. Monetary Policy

The nominal interest rate on bonds is determined by a monetary policy authority that sets it according to a feedback rule:

$$r_t = \bar{r} \left(\frac{\pi_t}{\pi} \right)^{\varphi_\pi} \left(\frac{y_t}{y_t^n} \right)^{\varphi_y} \exp\left(\epsilon_t^m\right) \ , \tag{9}$$

$$\epsilon_t^m = \rho_m \epsilon_{t-1}^m + \nu_t^m \quad \text{with} \quad \nu_t^m \overset{iid}{\sim} \left(0, \sigma_m^2\right) \ , \tag{10}$$

where $\bar{r}$, π and y_t^n are the steady state interest rate, aggregate inflation and the natural level of output (derived in the Internet Appendix), respectively. The parameter φ_π and φ_y denote the feedback coefficients that determine the sensitivity to deviations of inflation from its steady state and of output from its natural rate, respectively. Last, ϵ_t^m serves as monetary policy shock and evolves according to an AR(1)-process. We specify the monetary policy authority to use *current* inflation (opposed to its expectation), to avoid taking a stance on which type of expectations the monetary policy-maker has.

[10] Coibion and Gorodnichenko (2015) find firms' inflation expectations to be better captured by household survey data than by professional forecasters.

2.3. Equilibrium

Labour Market Equilibrium. As all intermediate firms produce with the same technology, equilibrium labour demand is symmetric. Aggregate working hours follow as:

$$l_t^d = \int_0^1 l_{i,t} di = \int_0^1 (y_{i,t})^{\frac{1}{\alpha}} di = (y_t)^{\frac{1}{\alpha}} \Delta_t^p = l_t^s = (1-\omega) \sum_{k=-\infty}^{t} \omega^{t-k} l_{t|k} \ , \tag{11}$$

where $\Delta_t^p \equiv \int_0^1 \left(\frac{p_{i,t}}{p_t}\right)^{-\frac{\varepsilon}{\alpha}} di$ is an index of relative price distortions. Since all firms face a symmetric maximisation problem, we focus on a symmetric price equilibrium so that $\Delta_t^p = 1$.

Goods Market Equilibrium. An equilibrium on the aggregate goods market requires that the total number of goods produced, y_t, equals the total amount of goods demanded, taking into account the dead-weight loss due to repricing cost:

$$y_t = c_t + \frac{\phi}{2}(\pi_t - 1)^2 y_t \ , \tag{12}$$

where $\pi_t = \frac{p_t}{p_{t-1}}$ denotes aggregate (gross) inflation.

Bond Market Equilibrium. Private bonds are in zero net supply, that is:

$$(1-\omega) \sum_{k=-\infty}^{t} \omega^{t-k} b_{t|k} = 0 \ . \tag{13}$$

New Keynesian Phillips Curve. Using the FOC on prices of intermediate good firms and symmetry, one can derive:

$$(\pi_t - 1)\, \pi_t = \omega \bar{E}_t \left[Q_{t,t+1} \frac{y_{t+1}}{y_t} (\pi_{t+1} - 1)\, \pi_{t+1} \right] + \frac{\varepsilon(1+\eta_l)}{\phi} (\mathrm{mc}_t^r - \mu) \ , \tag{14}$$

where mc_t^r are the real marginal cost, $\mu \equiv \frac{\varepsilon-1}{\varepsilon}$ denotes the steady state markup and $\eta_l = \frac{l}{1-l}$ denotes the stationary labour-leisure share. Note that *aggregate* inflation expectations, π_{t+1}, affect π_t. According to (14), optimal price setting requires inflation to be a function of current real marginal cost and expected future inflation.

Linearised Equilibrium Conditions. In the current formulation expectation heterogeneity matters for households' Euler equations and the Philips Curve. To arrive at an aggregated dynamic IS-curve, we follow the literature and rely on Branch and McGough (2009). First, we adopt their assumption on higher-order beliefs: household i's expectation about what another household k expects, is its own expectation:

$\tilde{E}_t^i \tilde{E}_t^k \mathbf{z}_{t+1} = \tilde{E}_t^i \mathbf{z}_{t+1}$, $i \neq k$ for some generic variable $\mathbf{z}$, which reduces the complexity imposed on the model considerably.[11] Second, we assume agents expect to hold the same wealth in the limit $t \to \infty$. For each agent i, consumption then equals the long-run consumption: $\tilde{E}_t^i (\hat{c}_\infty - \hat{c}_\infty^i) = 0$. This assumption prevents the wealth distribution from appearing in the aggregated IS-curve, which is beyond the scope of the paper.[12] After linearisation and aggregation we rewrite the model in terms of the output gap $\tilde{y}_t = \hat{y}_t - \hat{y}_t^n$ as a system of five equations and five variables $\{\tilde{y}_t, \hat{\pi}_t, \hat{r}_t, \hat{\epsilon}_t^m, u_t, \}_{t=0}^{\infty}$:

$$\tilde{y}_t = \bar{E}_t \tilde{y}_{t+1} - \left(\hat{r}_t - \bar{E}_t \hat{\pi}_{t+1}\right) , \tag{15a}$$

$$\hat{\pi}_t = \beta\omega \bar{E}_t \hat{\pi}_{t+1} + \kappa \tilde{y}_t + u_t , \tag{15b}$$

$$\hat{r}_t = \varphi_\pi \hat{\pi}_t + \varphi_y \tilde{y}_t + \hat{\epsilon}_t^m , \tag{15c}$$

$$\hat{\epsilon}_t^m = \rho_m \hat{\epsilon}_{t-1}^m + \nu_t^m , \tag{15d}$$

$$u_t = \rho_u u_{t-1} + \nu_t^u , \tag{15e}$$

where $\kappa \equiv \frac{\varepsilon(1+\eta_l)}{\alpha\phi}$ denotes the slope of the NK Phillips curve.[13] We introduce the cost-push shock u_t that could stem from a firm-specific shock to marginal cost to have a source of exogenous variation apart from the monetary policy innovation (see Ireland, 2004).[14] To solve the model, we need to specify how agents form expectations next.

3. Expectation Formation

In this section, we specify how different cohorts form expectations on inflation and the output gap based on Malmendier and Nagel (2016). We first explain how a single cohort forms expectations and then turn to why experience effects play an important role. Last, we highlight how the EBL-approach differs from CGL, which is among the most popular learning approaches.

3.1. Learning

Consider the model that agents use to form expectations. A large part of the literature assumes that agents know the true state-space representation (hence, they know the relevant variables for the economy's evolution) but have to learn about the coefficients of this

[11] For approaches explicitly taking into account higher order beliefs consider Angeletos et al. (2018) or Farhi and Werning (2019).

[12] This can be seen when aggregating the linearised Euler equations (see Appendix A). This assumption cancels the differences in expected consumption at $t \to \infty$ that occur when aggregating expected consumption for households that have heterogeneous expectations and solve a dynamic problem.

[13] The PLMs' updating equations (18a) and (18b) are also part of the model.

[14] In the Internet Appendix we derive the natural level of output and the natural rate that are zero in our model.

representation (e.g. Milani, 2007). Instead, we assume that agents employ a misspecified forecasting rule; that is, in comparison to having all state variables as regressors, agents use only a subset of them or no states at all. We adapt the set-up in Malmendier and Nagel (2016) and specify agents' PLM as AR(1).[15] However, we assume that it does not include a constant so that agents know the true mean of the model. Consequently, the PLM of a generic variable $\mathbf{z}_t$ for a household in cohort k is given by:

$$\mathbf{z}_{t|k} = b^{\mathbf{z}}_{t-1,k}\mathbf{z}_{t-1} + \varepsilon^{\mathbf{z}}_{t|k} \ , \tag{16}$$

where $\varepsilon^{\mathbf{z}}_{t|k}$ is a disturbance term which is serially-uncorrelated with zero mean and constant variance, and $b^{\mathbf{z}}_{t-1|k}$ is the estimated parameter of household k at time $t-1$. In our model, agents form expectations on the output gap and inflation. Hence, the set of variables on which agents form expectations is given by $Y^f \equiv \{\tilde{y}, \hat{\pi}\}$ with $\mathbf{z}_k \in Y^f$.

A further crucial point is the amount of information individuals are able to incorporate when forming expectations. We simplify the learning process and assume individuals that form expectations at time t use only information available at time $t-1$. By doing so, we avoid a simultaneity problem that arises when agents use time t endogenous variables to forecast future realisations, which in turn affects the time t endogenous variables. Hence, the realisation of time t endogenous variables and the formation of expectations of time $t+1$ variables are not simultaneously determined.[16]

The last element of the PLM is how its coefficients develop over time. Let $\mathbf{I}_t$ be the information set on which households base their forecast at time t. As explained above the information set $\mathbf{I}_t$ includes all model variables up to $t-1$. Consequently, the formation of expectations occurs before the realisation of the endogenous variables included in Y^f_t such that $\tilde{E}^k_t(\mathbf{z}_t) = \tilde{E}^k\left(\mathbf{z}_t|\mathbf{I}_t\right) \neq \mathbf{z}_t$. Instead, using (16), and presuming that the law of iterated expectations holds for the subjective expectations, households in cohort k forecast:

$$\begin{aligned}\tilde{E}^k_t\left(\mathbf{z}_{t+1}\right) = \tilde{E}^k_t\left(b^{\mathbf{z}}_{t|k}\mathbf{z}_t\right) &= \tilde{E}^k_t\left(b^{\mathbf{z}}_{t-1|k}\mathbf{z}_t\right) \\ &= \tilde{E}^k_t\left(b^{\mathbf{z}}_{t-1|k}\left(b^{\mathbf{z}}_{t-1,k}\mathbf{z}_{t-1} + \varepsilon^{\mathbf{z}}_{t|k}\right)\right) = (b^{\mathbf{z}}_{t-1|k})^2\mathbf{z}_{t-1} \ ,\end{aligned} \tag{17}$$

where for the first equality we use the PLM (dated $t+1$) and for the second that point estimates of the PLM parameters only include information up to $t-1$ (see Evans et al., 2013). The third equality makes use of the fact that agents form expectations before the current realisation of $\mathbf{z}$ such that also today's realisation is forecasted using the PLM. Finally, the last equality uses the fact that the PLM parameter estimated with information

[15] Orphanides and Williams (2005) and Slobodyan and Wouters (2012a) use similar specifications of agents' forecasting rule. The choice of order one is further consistent with a model under RE where variables are also Markov-processes of order one.

[16] This approach is also consistent with the literature (e.g. Evans and Honkapohja, 2012).

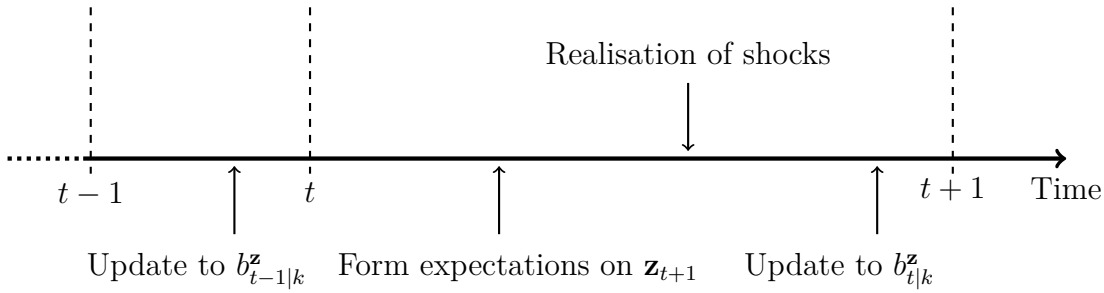

FIGURE 2: Timing Assumption of Updating the PLM Parameters

up to time $t-1$ is uncorrelated with the error term at time t, i.e. $\tilde{E}^k_t\left(b^{\mathbf{z}}_{t-1|k}\varepsilon^{\mathbf{z}}_{t|k}\right)=0$. After the realisation of time t shocks, agents update their PLM parameters from $b^{\mathbf{z}}_{t-1|k}$ to $b^{\mathbf{z}}_{t|k}$ using the following recursive least-squares (RLS, henceforth) algorithm:

$$b^{\mathbf{z}}_{t|k} = b^{\mathbf{z}}_{t-1|k} + \gamma_{t|k}\left(R^{\mathbf{z}}_{t|k}\right)^{-1}\mathbf{z}_{t-1}\hat{\varepsilon}^{\mathbf{z}}_{t|k} \tag{18a}$$

$$R^{\mathbf{z}}_{t|k} = R^{\mathbf{z}}_{t-1|k} + \gamma_{t|k}(\mathbf{z}_{t-1}\mathbf{z}'_{t-1} - R^{\mathbf{z}}_{t-1|k})\ , \tag{18b}$$

for each $\mathbf{z}\in Y^f$. Here, $\hat{\varepsilon}^{\mathbf{z}}_{t|k}\equiv\mathbf{z}_t - b^{\mathbf{z}'}_{t-1|k}\mathbf{z}_{t-1}$ denotes the forecast error of cohort k and $\gamma_{t|k}$ gives the (potentially) age-dependent Kalman gain of cohort k that assigns the relative importance of $\hat{\varepsilon}^{\mathbf{z}}_{t|k}$ with respect to the previous estimate, $b^{\mathbf{z}}_{t-1|k}$ and $R^{\mathbf{z}}_{t-1|k}$ (with $R^{\mathbf{z}}_{t-1|k}$ being the covariance matrix of estimates).

The lower the noise in the explanatory variables, the stronger the update. Importantly, we assume newly-born agents are endowed with a PLM parameter that is equal to the aggregate persistence parameter of the previous period.[17] We summarise the timing assumption in Figure 2.

3.2. Experience-Based Learning

The novelty of EBL lies in the age-dependent form of parameter updating. Concerning this aspect we deviate from the CGL-literature that sets the gain $\gamma_{t|k}$ equal to a constant g. Under CGL all agents update equally such that there is no heterogeneity in their PLM parameters. However, Malmendier and Nagel (2016) provide evidence that the gain parameter $\gamma_{t|k}$ depends on the amount of lifetime data (or equivalently age), $t-k$, of

[17] A newly-born agent follows the conventional wisdom. While several papers consider learning-from-experiences, the treatment of agents' initial belief varies: Schraeder (2015) uses initial beliefs that correspond to RE, Ehling et al. (2018) endow young agents with a small initial information set to deduct an initial belief and Collin-Dufresne et al. (2016) assume young agents inherit beliefs from their parents. In Appendix D we assume that each new-born cohort draws its initial parameter from a normal distribution around the RE estimate, where the normal distribution is truncated at ± 1. The key results remain unchanged.

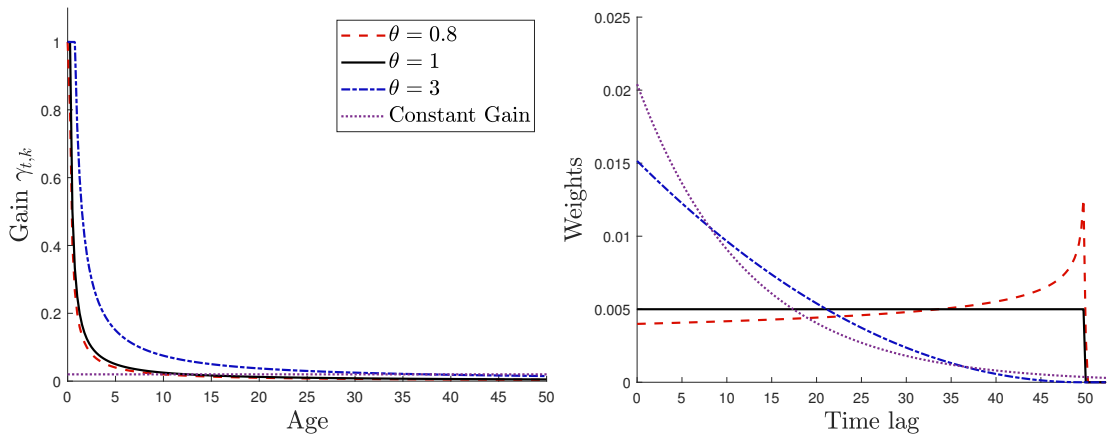

FIGURE 3: Gain and Weights on Past Data

Notes: The left panel denotes the evolution of the gain parameter over age (in years) for three different values of θ (the estimate of Malmendier and Nagel (2016) is around 3). The right panel shows how a 50 year old agents weights past information when estimating the parameters of the PLM (again for different θ). The purple line denotes the case of CGL.

individuals in cohort k,

$$\gamma_{t|k} = \begin{cases} \frac{\theta}{t-k} & \text{if } t-k \geq \theta \\ 1 & \text{if } t-k < \theta \end{cases}, \tag{19}$$

where $\theta > 0$ determines the degree to which individuals react to recent observations. Above specification implies, firstly, that expectations are heterogeneous between cohorts. Secondly, it implies that young agents have higher gains than older ones so that they update their PLM's parameters more strongly compared to their older peers.

Both aspects are captured in Figure 3. The left panel plots the gain parameter over age for different values of θ.[18] Young agents have high gains, consistent with the idea that they have less lifetime observations and, therefore, rely more on current data. The size of gains also decreases in age; the more so, the higher θ, which reflects that less weight is given to the more distant past. This is also captured in the right panel of Figure 3. It shows the implied weights a 50 year (200 quarter) old individual puts on data observed over its lifetime for different values of θ.[19] For $\theta > 1$, data observed early in life receives negligible weights as an individual ages so that recent data is more important to update the PLM (data before birth has weight zero - only lifetime information is used). Both panels taken together imply that, at time t, agents of different age estimate the PLM's parameters differently, which results in heterogeneity in expectations across cohorts. Note also that

[18]The graph is based on Malmendier and Nagel (2016). The Internet Appendix shows how to derive Figure 3.

[19]Note that RLS is the recursive formulation of weighted least squares (WLS, henceforth). The weights inside the weighting matrix contain the gain parameter $\gamma_{t|k}$ and, thus, depend on θ.

agents of different age use a different amount of information.[20] Although in our perpetual youth structure there may be individuals who use information from the far past, the mass of such a cohort declines as time passes by. Further, the weight such an individual would put on this information would be very small, such that this informations' influence on the current aggregate expectation is negligible.

3.3. Constant-Gain Learning

A feature of EBL is that agents have different gain parameters depending on their age. As mentioned above, studies of DSGE models with dynamic non-rational expectations instead often employ learning algorithms with CGL so that all agents react equally to new observations. In practice this amounts to replacing $\gamma_{t|k}$ in (18) with a constant, g, such that the left panel of Figure 3 shows a constant (purple line) across ages. Under this assumption, agents of different cohorts are homogeneous with respect to their expectation formation, i.e.

$$\tilde{E}_t^k \mathbf{z}_{t+1} = \tilde{E}_t \mathbf{z}_{t+1} = (b_{t-1}^{\mathbf{z}})^2 \mathbf{z}_{t-1} \ ,$$

for all cohorts k. This setup still retains the feature that new observations (and, hence, forecast errors) are weighted higher than old observations (right panel of Figure 3). However, each cohort weights them equally. In the following we will interpret the CGL approach as the counterpart of EBL where we *shut off* the effect of experiences on individuals' expectations. We simulate our model for this specification to study the *additional* endogenous source of variation that stems from experience effects alone.

4. Quantitative Analysis

This section presents a quantitative analysis of the model with EBL. We start with a brief description of our parameter choices. Next, we simulate the model under EBL and discuss its basic implications. We compare these to a model in which agents update their parameter estimates with a constant gain, which yields a key result we use for the subsequent analysis. To shed light on the impact on monetary policy transmission, we investigate impulse responses to monetary policy shocks under EBL, which we contrast with those from models with RE and CGL. We also explore how different relative sizes of young to old cohorts (we call this "demography", henceforth) affect results. Finally, we compare the monetary policy trade-off between output gap and inflation stabilisation in a

[20] Strictly speaking also newly entering agents would have access to all observations. However, as seen in the right panel of Figure 3, individuals do not put any weight on observations before birth. This stands in contrast to the SL-literature in which individuals weigh *all* data points, no matter how old they are.

Table 1: Parameter Choices (quarterly)

Variable		Value	
β	Discount factor	0.995	Ann. riskless rate 4%
ψ_n	Utility weight on leisure	1.17	Steady state labour supply of 1/3
ε	Elasticity of substitution	9	Mark-up of 12.5%
α	DRS parameter	0.66	U.S. labor share
ξ	(Inverse) Frisch elasticity	2	Standard choice
ϕ	Rotemberg parameter	91.9	Share non-adjusters 75%
φ_π	Taylor parameter π	1.5	Galí (2015)
φ_y	Taylor parameter y	0.125	Galí (2015)
π	Inflation Target	1	Zero-inflation steady state
ω	Survival probability	0.995	50 year working-life
ρ_u	Persistence $\hat{\epsilon}^u$	0.96	Ireland (2004)
ρ_m	Persistence $\hat{\epsilon}^m$	0.50	Galí (2015)
σ_u	Standard deviation ν^u (in %)	0.15	Ireland (2004)
θ	EBL parameter	3.044	Malmendier and Nagel (2016)
g	Gain under CGL	0.02	Milani (2007)

model with EBL to models with RE and CGL and, again, consider the role of demography.

4.1. Parametrisation and Solution Method

Parametrisation. One period in the model corresponds to one quarter. We calibrate the model's deep parameters to U.S. data (Table 1 gives a summary). Our choice of the survival probability $\omega = 0.995$ is guided to meet an average life span of 200 quarters, which represents the working-life of an agent.[21] Most of the other parameters are taken from the textbook model of Galí (2015). The households' discount factor β is calibrated to get a steady state real *annualised* return on riskless bonds of about 4% given our choice for ω. Furthermore, we set the steady state elasticity of substitution to $\varepsilon = 9$, which implies a steady state mark-up of 12.5%. The parameter of the production function α is chosen to be 0.66 in line with the labor share in U.S. data. The Rotemberg adjustment cost parameter is chosen to match a fraction of non-adjusters of 0.75 in a model with Calvo price setting.[22] As is standard in the literature, we set $\varphi_\pi = 1.5$ and $\varphi_y = 0.125$. Further, the values for the serial correlation coefficient of the cost-push shock, ρ_u, and the standard deviation of the innovation σ_u are set 0.96 and 0.0015, respectively, which correspond to the values estimated in Ireland (2004). Further, we follow Galí (2015) and set the serial correlation coefficient the monetary policy shock, ρ_m, to 0.5. When computing the impulse response functions to a monetary policy shock, the innovation to $\hat{\epsilon}_t^m$ corresponds to an increase of 25 basis points. We choose the learning parameter that

[21] Agents in our model are workers. We do not model retirement.

[22] If the fraction of non-adjusters is 0.75 (= average price duration of 4 quarters), $\phi = \frac{(\varepsilon-1)0.75}{(1-0.75)(1-0.75\beta\omega)}$.

governs the age-dependent gain under EBL as $\theta = 3.044$ (Malmendier and Nagel, 2016) and the CGL parameter g as 0.02 according to Milani (2007) and much of the learning literature. Finally, steady state inflation is targeted to be one while the time devoted to labour in steady state is 1/3.

Solution Method. We simulate the model under (i) EBL, (ii), RE, and (iii) CGL for the same random sequence of supply shocks, while setting the monetary policy innovation to zero. To initialise the PLM parameters for the learning models, we simulate the economy under RE for $T_{\text{init}} = 120$ quarters and estimate an AR(1) model for inflation and the output gap. The estimated AR(1) coefficients for both variables serve as the respective initial PLM parameter for the models with EBL and CGL. To start model simulations, we endow each cohort with the same initial persistence parameter, $b^{\mathbf{z}}_{-1|k} = b^{\mathbf{z}}_{-1}$, and the same covariance matrix of estimates $R^{\mathbf{z}}_{-1|k} = R^{\mathbf{z}}_{-1}$ for $\mathbf{z} \in Y^f$ both of which are updated subsequently. We then simulate the economy for $T_{\text{sim}} = T_b + 10,000$ quarters, where $T_b = 300$ is the number of periods that are discarded to wash out the impact of the initial values from the simulation of the RE economy. Each period, a member of cohort k updates its parameter estimate and the covariance matrix according to equation (18). Similar to Slobodyan and Wouters (2012a), we restrict agents to rely on covariance stationary forecasting models: we invoke a projection facility and restrict the new estimate to induce a stationary AR(1) process which requires $|b^{\mathbf{z}}_{t|k}| < 1$ for all cohorts k and $\mathbf{z} \in Y^f$.[23] Evans and Honkapohja (2012) argue that agents want to avoid explosive paths of the economy such that the agent chooses its parameter estimate accordingly. Each new born cohort's initial parameter equals the aggregate persistence parameter of the previous period. Further, since there is a infinite number of cohorts, we need to choose a finite number of cohorts for our simulation exercise. A high number of cohorts reduces the approximation error but is subject to the curse of dimensionality. Since the baseline model calibrates the survival probability such that the expected lifetime is 200 quarters, we restrict the number of cohorts in the aggregation to be 200 and normalise cohorts weights to sum to one. A detailed description of the algorithm is found in Appendix C.

4.2. Expectation Heterogeneity and the Experience Channel

In this section we discuss the key implications of EBL that matter for the analysis of monetary policy. In particular, young individuals' perceived persistence is, on average, lower and more volatile relative to old individuals. As a consequence under EBL, the *aggregate* perceived persistence in both inflation and the output gap is *on average* lower

[23] The restriction is invoked in only 6% of updates for PLM parameters under EBL. This falls to 2% when we marginally increase the bound on parameter estimates. Its usage is then similar to the one in Slobodyan and Wouters (2012a).

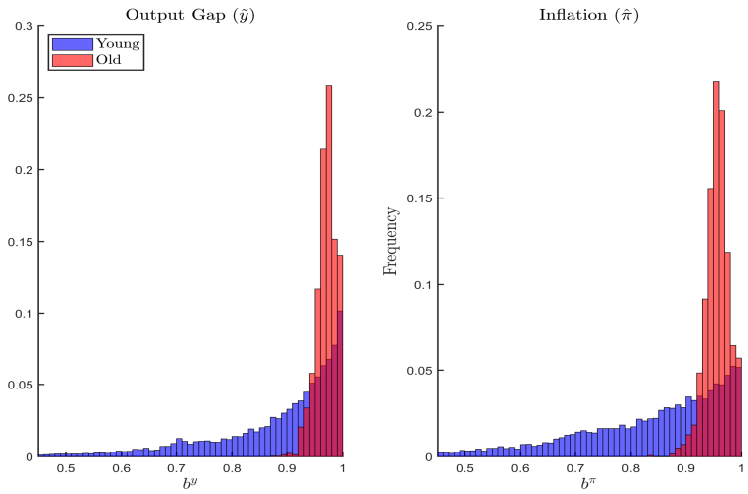

FIGURE 4: Distribution of PLM Parameters in the Model with EBL

Notes: The blue histogram denotes the distribution of PLM parameters for young ($k = 10$) individuals while the red histogram shows the distribution of the PLM parameters for old ($k = 158$). The left part shows the PLM parameter for the output gap while the right part shows the PLM parameter for inflation. We simulate the economy for $10,000$ periods.

compared to an economy where we shut-off experience effects, i.e. assuming CGL. This result follows from the effect of taking into account also very young agents on aggregate expectations.

Perceived Persistence. As pointed out in Malmendier and Nagel (2016), an important feature of EBL is that young agents' beliefs are more sensitive towards new observations compared to those of older individuals. In our model, the source of heterogeneity in expectations across cohorts stems from the heterogeneity in individuals' perceived persistence of both inflation and the output gap. In Figure 4 we plot the ergodic distribution of PLM parameters for individuals of cohort $k = 10$ and $k = 158$ for $10,000$ simulated periods. The left part of Figure 4 shows the distribution of PLM parameters for the output gap, while the right part shows the distribution of PLM parameters for inflation.
Figure 4 demonstrates that the perceived persistence in both the output gap and inflation of young agents is *on average* more dispersed and smaller relative to the one of old agents. There are two key driving forces that explain this finding. First, recall that young individuals rely on a lower amount of information when updating their PLM parameters and are more sensitive towards new observations. Since the variance in parameter estimates decreases in the number of observations used for updating and since young agents attach a high weight to new observations when updating, the dispersion of estimates is higher compared to the one of old agents. Second, young agents' over-weighting of the most recent observations and their limited amount of data also imply a lower perceived persistence, on average. Unless inflation or the output gap stay almost constant for some

FIGURE 5: Actual Inflation and PLM Parameters of Young and Old

Notes: The blue line depicts the annualised realised inflation deviation from steady state under EBL. The red lines show the estimated slope of the PLM for young (dashed) and old (dotted) cohorts. Throuhgout the simualtion we only compare agents of ages 10 and 158. Hence, we do not follow two single age groups through time.

periods, recurrent reversals make young agents perceive both variables to be less persistent than old individuals. Moreover, recall that we restrict agents to rely on covariance stationary forecasting models, as in Orphanides and Williams (2007) and Slobodyan and Wouters (2012b). Intuitively, this assumption ensures that agents reject explosive forecasting models with a PLM parameter above one. The combination of a higher dispersion in young individuals' beliefs and the truncation of the PLM parameter distribution further decreases the mean of young individuals' beliefs.

Intuition Behind the Learning Process. The dynamics of the difference in expectations is linked to the experienced inflation of each cohort. Figure 5 displays a snapshot of one simulation path. The blue thick line depicts simulated *realised* inflation (left y-axis) and the red lines denote the slope of the PLM for a young (dashed) and an old (dotted) cohort (right y-axis). Two aspects stand out. First, movements in *actual* inflation are reflected in PLM parameters with a lag which is a consequence of the timing assumption. Recall that the update step only uses information up to the last period, so that any shift in inflation is included with a one-period lag. Second, PLM parameters of young and old households differ, which, in turn, determines differences in expected inflation. A key reason for the difference in the perceived persistence is the fact that young individuals only use the most recent data to estimate the parameter of their forecasting model. Consider the region around period $t = 75$. During the preceding ten quarters, inflation stayed roughly at the same value so that young individuals perceived inflation to be very persistent. In fact, their perceived persistence was higher than the one of old individuals who

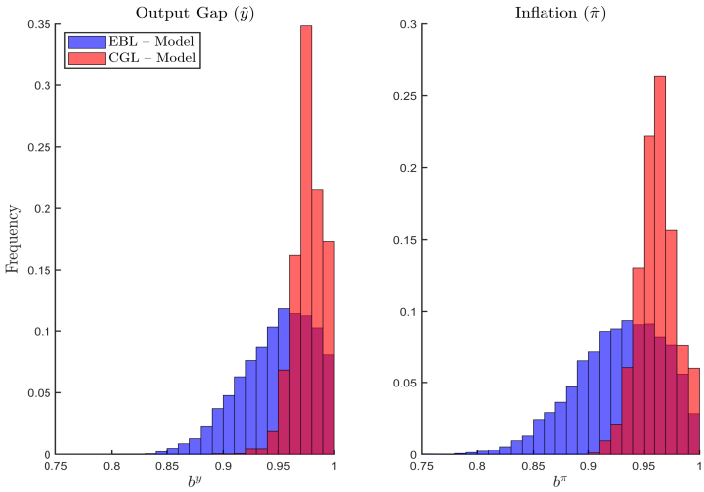

FIGURE 6: Distribution of PLM Parameters in the Models with EBL and CGL

Notes: The blue histogram denotes the distribution of PLM parameters in the model under EBL while the red histogram shows the distribution of the PLM parameters under CGL. The left part shows the PLM parameter for the output gap while the right part shows the PLM parameter for inflation. We simulate the economy with EBL and CGL for 10,000 periods.

already experienced sharp reversals in the inflation. In contrast, around quarter $t = 30$, inflation has gone through recurrent reversals during the preceding ten quarters so that young individuals perceive inflation to be less persistent than old individuals who also observed more stable times of inflation. Figure 5 shows that young individuals, on average, perceive inflation to be less persistent which is consistent with the results in Figure 4.

Endogenous Reduction in the Aggregate Perceived Persistence. The heterogeneity in the perceived persistence across cohorts has important implications for the *aggregate* perceived persistence, which, under EBL, is a size-weighted average over cohort-specific perceived persistences. In Figure 6 we compare the distribution of the aggregate perceived persistence under EBL (blue histogram) to the one obtained under CGL (red histogram). Under EBL, the aggregate perceived persistence for both inflation (right panel) and the output gap (left panel) is lower, *on average*, and more dispersed. To understand this finding, note that for the calibration in Table 1, the gain of the representative agent under CGL maps into the gain of an individual with a 38 year working life under EBL. For any individual in the model with EBL that is younger than the representative agent under CGL, the perceived persistence of both inflation and the output-gap is lower, on average, and more dispersed. If the mass of individuals younger than the representative agent under CGL is sufficiently high, also the aggregate perceived persistence is lower, on average, and more dispersed. Hence, the age distribution affects the first two moments of the distribution of the aggregate perceived persistence. EBL *endogenously* pushes down

the aggregate perceived persistence of inflation and the output gap, on average.[24] Put differently, CGL *overstates* the reliance on past data when forming expectations compared to EBL. As we show in the next section, this is of relevance for the impact of monetary policy on expectations and its transmission on inflation.

Experience Channel. Since the aggregate perceived persistence is a size-weighted average over cohort-specific perceived persistences, a variation in the age distribution directly affects the aggregate perceived persistence through a composition effect, which we call the *Experience Channel*. In our model, a variation in the age distribution corresponds to a variation in the survival probability, ω. As ω decreases, the share of young individuals increases while the share of old individuals decreases. As shown above, young individuals' perceived persistence is on average lower than the one of old ones so that the aggregate perceived persistence for inflation and the output gap decreases in the share of young individuals. In turn, the aggregate perceived persistence for these two variables under CGL is hardly affected. Thereby, EBL opens up a new channel by which a variation in the age distribution affects the transmission of monetary policy.

4.3. The Effect of Monetary Policy Under EBL

We study the effect of EBL on monetary policy and relate the magnitude of this effect to the age-distribution. We show that the pass-through of monetary policy on aggregate variables via expectations under EBL is weakened relative to models with RE or CGL. The attenuation of the pass-through of monetary policy affects both the transmission of monetary policy shocks as well as its trade-off between stabilising inflation and output gap closure under supply shocks. We also compare economies with a high share of old cohorts against an economy with many young cohorts if EBL is used. In older economies the Experience Channel leads to stronger and more persistent responses to monetary policy shocks and aggravates the stabilisation trade-off. Thus, models that abstract from expectation heterogeneity based on experience effects miss important factors that determine the degree of monetary policy transmission.

4.3.1. Transmission of a Monetary Policy Shock

First, we analyse the effect of EBL on the transmission of monetary policy (MP) shocks and how this effect depends on the age distribution. To explore this we compute *generalised* impulse response functions in the model with (i) EBL, (ii) CGL, and (iii) RE. The

[24] In Appendix D, we perform a robustness check by varying key parameters of the model. The core result regarding the aggregate perceived persistence under EBL remains unchanged.

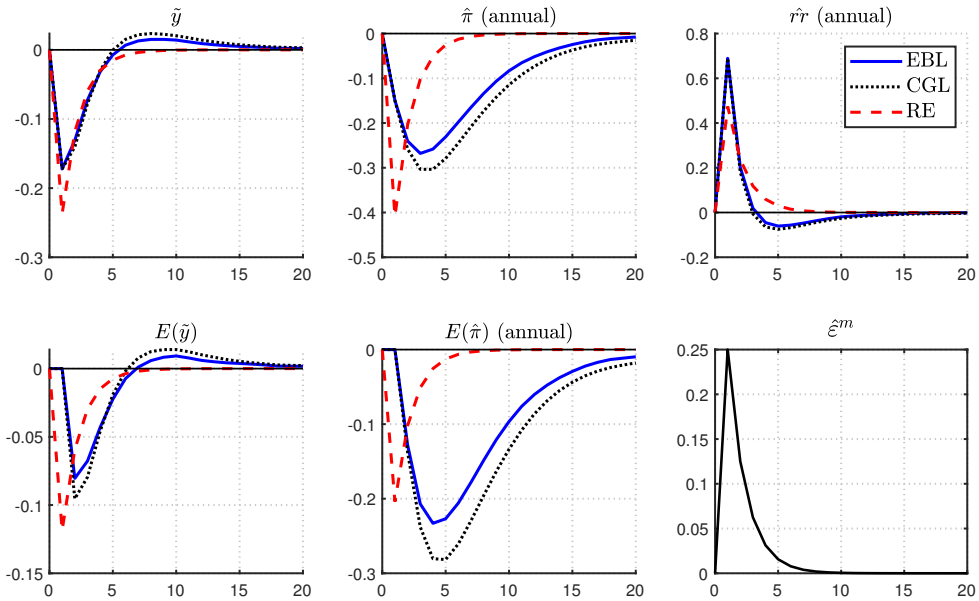

FIGURE 7: IRFs to a MP Shock

Notes: We show the IRFs for key variables in the economies under RE (red), CGL (black) and EBL (blue). We average responses over 8,000 iterations to generate mean impulse responses. The output gap, output gap expectations as well as the monetary policy shock are measured as percentage deviations from their respective steady state while the other variables are measured as (annualised) deviation from their respective steady state.

monetary policy shock corresponds to a innovation of 25 basis points to $\hat{\epsilon}_t^m$. The algorithm used to compute the generalised impulse response functions is described in Appendix C.

Impact. Consider Figure 7. On impact, a contractionary MP shock affects both the output gap and inflation negatively. In response, the central bank pushes the nominal interest rate down. This decrease is, however, not sufficient to offset the exogenous shift such that the real interest rate increases, on impact. Comparing the initial responses, both the output gap and inflation react less under EBL compared to the RE model. If we shut off experience effects on expectations (CGL), we observe similar impact responses of the output gap and of inflation that become stronger the period after the shock hits the economy when expectations are revised.

Revision of Expectations. Under both EBL and CGL, individuals revise their beliefs with a delay of one period. This backward-looking expectation formation is visualised in the lower half of Figure 7, which illustrates responses of expectations. In the period *after* the shock, individuals revise beliefs on future inflation and the output gap downwards, which feeds back into the current output gap and inflation and creates hump-shaped responses in Figure 7. The lower impact response of inflation and, to a smaller extent, the output gap under EBL compared to CGL results from the smaller reaction of expectations that, in turn, is driven by a smaller aggregate perceived persistence attached to these

two variables (see Figure 6). Under adaptive expectations the perceived persistence is a result of past shocks and monetary policy actions. Since agents under EBL attach, on average, a lower perceived persistence to variables also monetary policy is less able to affect expectations and thereby contemporaneous variables. As a result the pass-through of monetary policy on inflation is impaired when we assume EBL. The difference in the response of the output gap under EBL and CGL is small because the simultaneous decrease in the real rate (which ceteris paribus increases demand) partly countervails the impact of more negative CGL-expectations.

Forward-looking RE stand in contrast to adaptive expectations. Agents take into account the Taylor rule's impact on the future path of real interest rates as well as the shock's persistence when forming expectations. As a result, RE of the output gap react stronger than those based on CGL or EBL because the persistence of the shock increases the nominal rate also for future periods, which implies a smaller output gap so that the expected output gap is lower. As consequence, also expected inflation is less negative compared to the economies with adaptive expectations in which agents are incapable of perceiving these changes due to their backward-looking expectations.

Dynamics. The dynamic response of macroeconomic variables under EBL (or CGL) is quite different compared to the one under RE. While under RE the economy reverts back to the steady state roughly after ten quarters, deviations in the economy under EBL are more persistent. This slower reversion to the steady state results from the weaker pass-through of monetary policy which prolongs the effect of the MP shock compared to the economy under RE. Moreover, under EBL, agents revise their beliefs downwards and the shock is more slowly transmitted into their expectations compared to RE where individuals take into account the actual data generating process and perfectly internalise the shock process into their expectations. In addition, inflation displays a hump-shaped response under EBL which results from the backward-looking expectations of agents and their lagged response to the monetary policy shock. In turn, the output gap does not display this hump-shaped behaviour but starts reversing to its steady state value directly after the initial response. Notwithstanding the reduced pass-through of monetary policy on aggregate demand, the decrease in the real interest rate is sufficient to reverse the drop in the output gap in the first period after the MP shock hits the economy.

To bring inflation back to its steady state, monetary policy is decreasing the nominal interest rate by that much that the real interest rate undershoots, which, in turn, results in an overshooting of the output gap that boosts inflation upwards towards its steady state. This overshooting is more pronounced if we shut off the experience effect so that the gap between the inflation responses of the models under EBL and CGL closes.

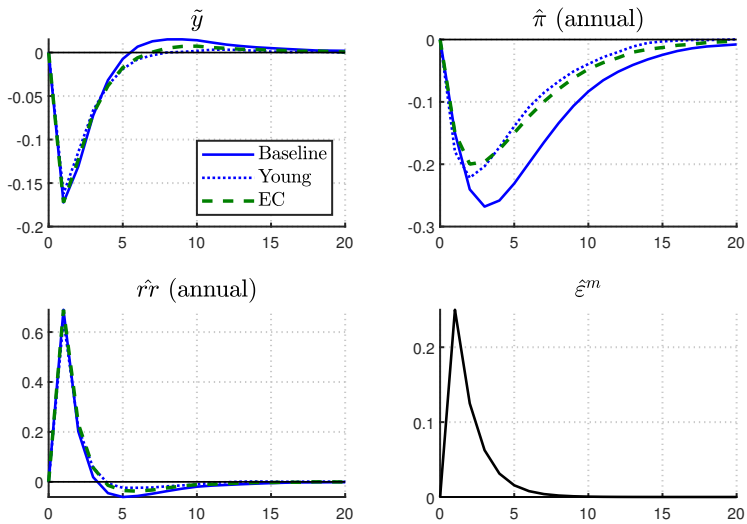

FIGURE 8: IRFs to a MP Shock: Decomposition Under EBL

Notes: We show the IRFs for key variables in economies under EBL. We distinguish the baseline case (blue solid), the young economy (red), and the isolated Experience Channel (green). We average responses over 8,000 iterations to generate mean impulse responses. The output gap and the monetary policy shock are measured as percentage deviations from their respective steady state while the other variables are measured as (annualised) deviation from their respective steady state.

Demography. Next, we focus on the impact of a demographic shift in the model with EBL and analyse the effect on the transmission of the monetary policy shock.[25] In the perpetual youth model, a change in the demographic structure corresponds to a change in the survival probability ω. To illustrate the effect of the demographic structure we reduce the survival probability to 0.9. A change in the survival probability affects,

1. the effective discount factor, $\tilde{\beta} \equiv \beta\omega$
2. the slope of the Phillips-Curve (via the change in $\tilde{\beta}$)
3. *under EBL*, aggregate expectations, $\bar{E}_t$.

For the sake of the argument, we shut-off channels 1. and 2. by varying β such that $\tilde{\beta}$ stays constant and only consider the variation that stems from channel 3.. In Figure 8, the solid blue line denotes responses in the baseline ($\omega = 0.995$) economy and the green line shows the impulse response function under EBL when the survival probability is lowered to 0.9 and only affects the economy through the Experience Channel (channel 3.). In contrast the dotted blue line denotes responses when all channels are active.
As discussed in Section 4.2, an increase in the share of young agents (i.e. decrease in ω) reduces the aggregate perceived persistence of both inflation and the output gap via the Experience Channel. Thus, in the young economy (blue dotted line) the pass-through of monetary policy is reduced. The lower aggregate perceived persistence reduces the impact

[25] In the Internet Appendix we perform the same exercise for the models under RE and CGL. Given no Experience Channel exists, differences between a young and an old economy are small.

response of the output gap and inflation. Further, the response of both variables displays less persistence compared to the baseline economy as the aggregate *perceived* persistence is lower.

The Experience Channel alone generates a considerable decrease in the transmission of monetary policy on inflation. In particular, the additional variation in the transmission of monetary policy when taking into account all three channels 1.-3. is relatively small. When taking into account all channels, the impact response is slightly more pronounced compared to the response that accounts only for channel 3.. This stems from the higher slope of the NKPC which increases the sensitivity of inflation to movements in aggregate demand. However, the persistence of the response is lower because (i) current inflation is less sensitive to inflation expectations because $\tilde{\beta}$ decreases, and (ii) the transmission of monetary policy is higher through the increase in the slope of the NKPC.

4.3.2. Trade–off Under Supply Shocks

Next, we compare different Taylor rule calibrations with respect to their ability to close the output gap and to stabilise inflation under supply shocks. In the context of the NK model, it is a well-known result that the Taylor rule is incapable to close the output gap and to stabilise inflation at the same time when supply shocks perturb the economy (e.g. Galí, 2015). Since EBL impairs the monetary policy transmission, the trade-off may also be affected. Starting from the baseline parametrisation of the Taylor parameters (φ_π, φ_y), we successively increase monetary policy's output gap stabilisation motive (increase φ_y up to 1), while holding constant the Taylor parameter on inflation ($\varphi_\pi = 1.5$). The results are shown in Figure 9 where we plot the standard deviation of the output gap (σ_y) against the one of inflation (σ_π) for each combination of (φ_π, φ_y). The blue line displays the results under EBL, the black line shows corresponding results when we shut off the experience effects (CGL) and the red line displays the simulation results under RE.

On the one hand, the monetary policy trade-off between closing the output gap and stabilising inflation under supply shocks prevails under EBL. As we increase the Taylor coefficient on the output gap, its volatility decreases, while inflation volatility increases. On the other hand, the policy frontier shifts inwards under EBL which implies that for each combination of the Taylor coefficients, the output gap is *less* volatile. In contrast to CGL, the slope of the policy frontier is also flatter, i.e. the trade-off attenuates.

The inward shift of the EBL policy frontier is explained by two effects. First, the pass-through of monetary policy is weaker under adaptive expectations *in general*. RE are forward-looking and consider the Taylor rule's impact on the future path of real interest rates, which directly affects the current output gap via the Dynamic IS Curve. Under EBL (and CGL), in contrast, agents' expectations are backward-looking. They fail to project the effect of monetary policy on the future path of the real interest rate. As a

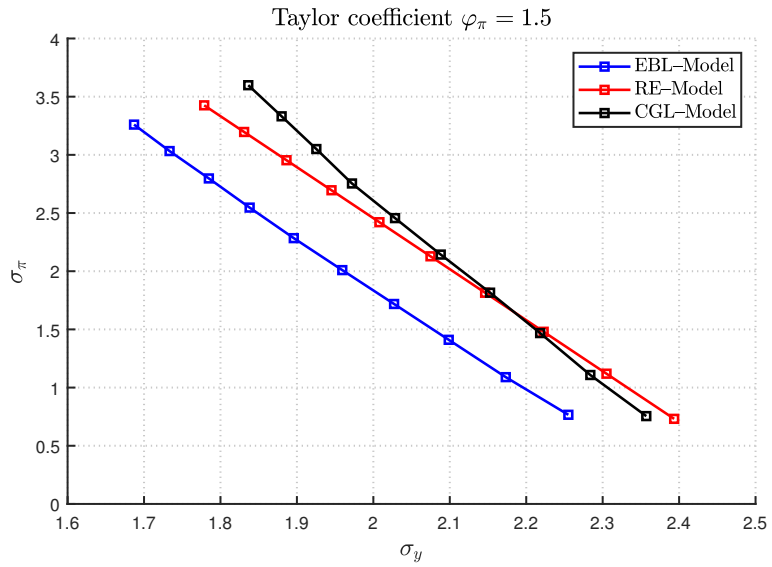

FIGURE 9: Monetary Policy Frontier Under a Supply Shock

Notes: We define a grid of points for φ_y. We then simulate the economy as described above for 10,000 periods for each grid point while holding $\varphi_\pi = 1.5$. We then plot the standard deviation of the output gap against the one for inflation for the models under EBL (blue), CGL (black) and with RE (red). Units are in percent deviation from the steady state.

consequence, the effect of the real interest rate on aggregate demand is subdued which dampens the effect of aggregate demand on inflation (relative to the model with RE).[26] Second, the aggregate perceived persistence on both, inflation and the output gap, is, *on average*, lower under EBL compared to CGL. In consequence, expectations, and thereby current variables, react by less to monetary policy or supply-side shocks so that output is less volatile and the policy frontier shifts inwards compared to the one under CGL.[27] The flatter policy frontier in comparison to CGL is also a result of the reduction in the aggregate perceived persistence. When increasing the output gap stabilisation motive, nominal interest rate changes more closely follow movements in the output gap and thereby stabilise it. However, inflation becomes more volatile ceteris paribus. This effect is amplified via the response of expectations that react to greater movements in current inflation. If, however, expectational responses are muted, as is the case under EBL, this effect is weaker so that for an increasing output gap stabilisation motive, the associated increase in inflation volatility is lower under EBL.

[26]Note that for an increasing focus on output gap stabilisation, we observe the CGL policy frontier to be above the one for RE. In these cases the output gap is *more* volatile despite adaptive expectations. While RE internalise that nominal rate movements via the simple interest rate rule now follow, and hence stabilise, the output gap, adaptive expectations do not. Monetary policy's weaker ability to influence backward-looking expectations then curtails the reduction in output gap volatility.

[27]The Internet Appendix performs the same analysis for a higher focus on inflation. Our results remain unchanged.

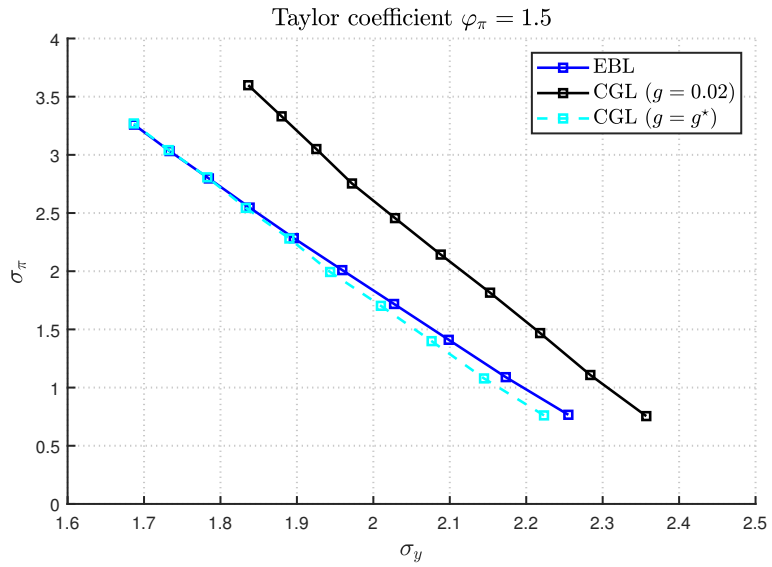

FIGURE 10: Monetary Policy Frontier Under a Supply Shock: High Constant Gain

Notes: The figure shows the policy frontier in the model under the baseline calibration under (a) EBL, (b) CGL when the gain parameter is set to $g = 0.02$; (c) CGL when the gain parameter is set such that the policy frontier under CGL comes close to the one obtained under EBL ($g^* = 0.12$).

EBL vs. CGL. In fact, it is possible to replicate our finding regarding the position of the policy frontier also with CGL. Figure 10 depicts policy frontiers for models with EBL and CGL. The solid blue line shows our EBL case in which the gain of agents, depending on age, ranges from 0.015 to 1. If in contrast, we choose a model based on CGL the literature (see estimates in Milani, 2007; Slobodyan and Wouters, 2012b) suggests a gain of $g = 0.02$ (black line). Only when setting the CGL parameter to $g^* = 0.12$ we can replicate the EBL frontier with CGL expectation formation (light blue). Yet, this does not invalidate the usage of EBL for two reasons. First, such a value for the gain parameter is empirically implausible.[28] Second, in such an economy the age distribution does not affect aggregate expectations. The model with EBL reveals that expectation heterogeneity matters for aggregate outcomes and hence monetary policy. To strengthen this point we, again, vary the demographic structure and discuss its effect on monetary policy via the Experience Channel next.

Demography. We consider how the policy frontier is affected by a change in the demographic structure of the economy by reducing ω to 0.9 (corresponding to a high share of young agents). Recall that an increase in the share of young agents further reduces the aggregate perceived persistence under EBL. As can be seen in Figure 11, this reduction in the aggregate PLM parameter shifts the policy frontier downwards (blue dotted). Hence, this demography-driven shift reinforces the Experience Channel. Moreover, the slope of

[28]For example, Milani (2007) finds that the 95% posterior probability interval for the constant gain parameter to be $[0.0133, 0.0231]$.

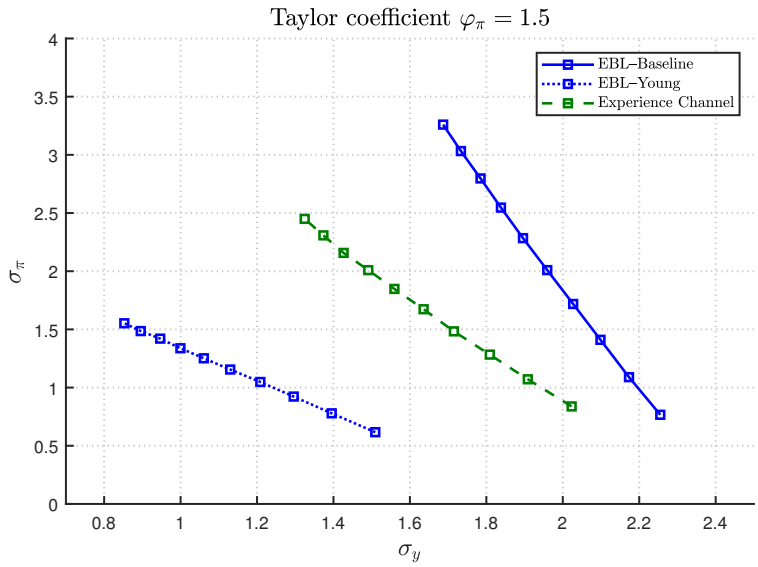

FIGURE 11: Monetary Policy Frontier Under a Supply Shock: Experience Channel

Notes: In addition to the remarks on Figure 9: we increase the share of young agents while holding the effective discount rate $\tilde{\beta} = \beta\omega$ constant. We also show the frontiers for the models under EBL in an old economy (blue solid) and in a young economy. The latter is divided into the full effect (blue dotted) and into the effect from experiences alone (green).

the policy frontier flattens which indicates that a higher weight on the Taylor coefficient on the output gap increases inflation volatility by less compared to the baseline model (blue solid). Thus, the trade-off becomes weaker in an economy with a high share of young individuals. Considering the effect of the Experience Channel alone (green line) we observe that it already generates a substantial shift of the policy frontier. Intuitively, in an economy with a higher share of young individuals, the perceived persistence in inflation is lower such that the effect of inflation expectations on current inflation is muted. This reduces the impact response of inflation and the lagged effect of shocks through inflation expectations on current inflation.[29] The total effect is more pronounced as the effective discount factor is decreasing (channel 1.), which attenuates the effect of inflation expectations on current inflation even further.[30]

5. Conclusion

This paper aims to answer the question how experience-based heterogeneity in expectations affects monetary policy's effectiveness in stabilising inflation and closing the output gap. To address this issue, we introduce a NK model with overlapping generations in which

[29] Conversely, inflation is more responsive in old economies. This is also observed empirically. For a panel of OECD countries Baksa and Munkácsi (2020) identify a significant and positive impact of a higher old-age dependency ratio on inflation volatility.

[30] In the Internet Appendix, we show that if we shut off the experience channel, the demography-driven shift in the policy frontier is substantially less pronounced and stays close to the policy frontier under RE.

individuals forecast inflation and the output gap based on a simplified model. Motivated by the empirical evidence of Malmendier and Nagel (2016), individuals' expectations depend on their respective lifetime experiences, which, due to the presence of differently aged cohorts, creates heterogeneity in expectations.

Expectations heterogeneity in our model is based on the heterogeneity in the perceived persistence across cohorts. We find that under EBL, the *aggregate* perceived persistence in the economy is pushed down relative to a model with CGL where agents attach the *same constant* weight to new information. Importantly, we show that a model with CGL cannot replicate the endogenous reduction in the aggregate perceived persistence under EBL unless the weight individuals attach to new information is set to an empirically implausible value.

We find that, under EBL, the pass-through of monetary policy turns weaker relative to models with RE or CGL. The demographic structure directly affects aggregate expectations, which are a size-weighted average across cohort, through a composition effect, which we call the *Experience Channel*. We demonstrate that the Experience Channel matters for monetary policy in particular as the share of old individuals increases and highlight two effects. First, the monetary policy trade-off between output gap closure and inflation stabilisation under supply shocks aggravates. Second, the response of inflation to a monetary policy shock is *both* more pronounced and more persistent. Thus, the age structure is a relevant factor to determine the transmission of monetary policy.

A worthwhile extension of our analysis would be to consider the effect of experience-based heterogeneity on the conduct of *optimal* monetary policy. Based on a second-order approximation of the welfare-function similar to Di Bartolomeo et al. (2016), one could derive welfare-maximizing monetary policy reaction functions. As shown by Di Bartolomeo et al. (2016), under heterogeneous expectations the welfare loss function entails consumption dispersion across households with different expectations. However, Gasteiger (2018) questions the practical relevance of the welfare loss function due to the issue that consumption heterogeneity driven by heterogeneous expectations is not readily observable in the data. However, given the assumption that individuals within the same cohort have homogenous expectations, the distribution of consumption across age groups is perfectly observable for policymakers. Hence, EBL might be beneficial for practical reasons when aiming to analyse optimal monetary policy under heterogeneous expectations.

References

Acedański, Jan (2017). Heterogeneous expectations and the distribution of wealth. *Journal of Macroeconomics* 53, pp. 162–175. DOI: 10.1016/j.jmacro.2017.06.009.

Adam, Klaus, Albert Marcet, and Juan P. Nicolini (2016). Stock market volatility and learning. *Journal of Finance* 71(1), pp. 33–82. DOI: 10.1111/jofi.12364.

Aksoy, Yunus, Henrique S. Basso, Ron P. Smith, and Tobias Grasl (2019). Demographic structure and macroeconomic trends. *American Economic Journal: Macroeconomics* 11(1), pp. 193–222. DOI: 10.1257/mac.20170114.

Angeletos, George-Marios, Fabrice Collard, and Harris Dellas (2018). Quantifying confidence. *Econometrica* 86(5), pp. 1689–1726. DOI: 10.3982/ecta13079.

Baksa, Dániel and Zsuzsa Munkácsi (2020). More Gray, More Volatile? Aging and (Optimal) Monetary Policy. *Dynare Working Papers* (58).

Berardi, Michele (2009). Monetary policy with heterogeneous and misspecified expectations. *Journal of Money, Credit and Banking* 41(1), pp. 79–100. DOI: 10.1111/j.1538-4616.2008.00188.x.

Berg, Kimberly, Chadwick Curtis, Steven Lugauer, and Nelson C. Mark (2020). Demographics and Monetary Policy Shocks. *Journal of Money, Credit, and Banking* forthcoming.

Blanchard, Olivier J. (1985). Debt, deficits, and finite horizons. *Journal of Political Economy* 93(2), pp. 223–247. DOI: 10.1086/261297.

Branch, William A. (2004). The theory of rationally heterogeneous expectations: evidence from survey data on inflation expectations. *The Economic Journal* 114(497), pp. 592–621. DOI: https://doi.org/10.1111/j.1468-0297.2004.00233.x.

Branch, William A. and Bruce McGough (2009). A New Keynesian model with heterogeneous expectations. *Journal of Economic Dynamics and Control* 33(5), pp. 1036–1051. DOI: 10.1016/j.jedc.2008.11.007.

— (2011). Business cycle amplification with heterogeneous expectations. *Economic Theory* 47(2-3), pp. 395–421. DOI: 10.1007/s00199-010-0541-2.

— (2018). Heterogeneous expectations and micro-foundations in macroeconomics. In: *Handbook of Computational Economics*. Vol. 4, pp. 3–62. DOI: 10.1016/bs.hescom.2018.03.001.

Carvalho, Carlos, Andrea Ferrero, and Fernanda Nechio (2016). Demographics and real interest rates: Inspecting the mechanism. *European Economic Review* 88, pp. 208–226. DOI: https://doi.org/10.1016/j.euroecorev.2016.04.002.

Coibion, Olivier and Yuriy Gorodnichenko (2015). Is the Phillips curve alive and well after all? Inflation expectations and the missing disinflation. *American Economic Journal: Macroeconomics* 7(1), pp. 197–232. DOI: 10.1257/mac.20130306.

Collin-Dufresne, Pierre, Michael Johannes, and Lars A. Lochstoer (2016). Asset pricing when 'this time is different'. *Review of Financial Studies* 30(2), pp. 505–535. DOI: 10.1093/rfs/hhw084.

De Grauwe, Paul (2011). Animal spirits and monetary policy. *Economic Theory* 47(2-3), pp. 423–457. DOI: 10.1007/s00199-010-0543-0.

Di Bartolomeo, Giovanni, Marco Di Pietro, and Bianca Giannini (2016). Optimal monetary policy in a New Keynesian model with heterogeneous expectations. *Journal of Economic Dynamics and Control* 73, pp. 373–387. DOI: 10.1016/j.jedc.2016.10.003.

Eggertsson, Gauti B., Neil R. Mehrotra, and Jacob A. Robbins (2019). A model of secular stagnation: Theory and quantitative evaluation. *American Economic Journal: Macroeconomics* 11(1), pp. 1–48. DOI: 10.1257/mac.20170367.

Ehling, Paul, Alessandro Graniero, and Christian Heyerdahl-Larsen (2018). Asset prices and portfolio choice with learning from experience. *Review of Economic Studies* 85(3), pp. 1752–1780. DOI: 10.1093/restud/rdx077.

Eusepi, Stefano and Bruce Preston (2018). The Science of Monetary Policy: An Imperfect Knowledge Perspective. *Journal of Economic Literature* 56(1), pp. 3–59. DOI: 10.1257/jel.20160889.

Evans, George W. and Seppo Honkapohja (2012). Learning and expectations in macroeconomics. Princeton University Press.

Evans, George W., Seppo Honkapohja, and Kaushik Mitra (2013). Notes on Agents' Behavioral Rules Under Adaptive Learning and Studies of Monetary Policy. In: *Macroeconomics at the Service of Public Policy*. Ed. by Thomas J. Sargent and Jouko Vilmunen. Oxford University Press. Chap. 4. DOI: 10.1093/acprof:oso/9780199666126.003.0005.

Farhi, Emmanuel and Iván Werning (2019). Monetary Policy, Bounded Rationality, and Incomplete Markets. *American Economic Review* 109(11), pp. 3887–3928. DOI: 10.1257/aer.20171400.

Galí, Jordi (2015). Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications. 2nd ed. Princeton: Princeton University Press.

— (2021). Monetary Policy and Bubbles in a New Keynesian Model with Overlapping Generations. *American Economic Journal: Macroeconomics* 13(2), pp. 121–67.

Gasteiger, Emanuel (2014). Heterogeneous expectations, optimal monetary policy, and the merit of policy inertia. *Journal of Money, Credit and Banking* 46(7), pp. 1535–1554. DOI: 10.1111/jmcb.12149.

— (2018). Optimal Constrained Interest-Rate Rules under Heterogeneous Expectations. *Working Paper*.

Hagenhoff, Tim (2018). An aggregate welfare maximizing interest rate rule under heterogeneous expectations. *BERG Working Paper Series* (139).

Ireland, Peter N (2004). Technology shocks in the new Keynesian model. *Review of Economics and Statistics* 86(4), pp. 923–936. DOI: https://doi.org/10.1162/0034653043125158.

Kara, Engin and Leopold von Thadden (2016). Interest Rate Effects of Demographic Changes in a New Keynesian Life-Cycle Framework. *Macroeconomic Dynamics* 20, pp. 120–164. DOI: 10.1017/S1365100514000273.

Kaustia, Markku and Samuli Knüpfer (2008). Do investors overweight personal experience? Evidence from IPO subscriptions. *Journal of Finance* 63(6), pp. 2679–2702. DOI: 10.1111/j.1540-6261.2008.01411.x.

Malmendier, Ulrike and Stefan Nagel (2011). Depression babies: do macroeconomic experiences affect risk taking? *Quarterly Journal of Economics* 126(1), pp. 373–416. DOI: 10.1093/qje/qjq004.

— (2016). Learning from inflation experiences. *Quarterly Journal of Economics* 131(1), pp. 53–87. DOI: 10.1093/qje/qjv037.Advance.

Malmendier, Ulrike, Stefan Nagel, and Zhen Yan (2020a). The making of hawks and doves. *Journal of Monetary Economics* forthcoming. DOI: https://doi.org/10.1016/j.jmoneco.2020.04.002.

Malmendier, Ulrike, Demian Pouzo, and Victoria Vanasco (2020b). Investor experiences and financial market dynamics. *Journal of Financial Economics* 136(3), pp. 597–622. DOI: 10.1016/j.jfineco.2019.11.002.

Mankiw, Gregory N. and Ricardo Reis (2007). Sticky information in general equilibrium. *Journal of the European Economic Association* 5(2-3), pp. 603–613. DOI: 10.1162/jeea.2007.5.2-3.603.

Massaro, Domenico (2013). Heterogeneous expectations in monetary DSGE models. *Journal of Economic Dynamics and Control* 37(3), pp. 680–692. DOI: https://doi.org/10.1016/j.jedc.2012.11.001.

Milani, Fabio (2007). Expectations, learning and macroeconomic persistence. *Journal of Monetary Economics* 54(7), pp. 2065–2082. DOI: 10.1016/j.jmoneco.2006.11.007.

Nagel, Stefan and Zhengyang Xu (2019). Asset pricing with fading memory. *NBER Working Paper No. w26255*. National Bureau of Economic Research. DOI: 10.3386/w26255.

Orphanides, Athanasios and John Williams (2005). The decline of activist stabilization policy: Natural rate misperceptions, learning, and expectations. *Journal of Economic Dynamics and Control* 29(11), pp. 1927–1950. DOI: 10.1016/j.jedc.2005.06.004.

— (2007). Robust monetary policy with imperfect knowledge. *Journal of Monetary Economics* 54(5), pp. 1406–1435. DOI: https://doi.org/10.1016/j.jmoneco.2007.06.005.

Pfajfar, Damjan and Emiliano Santoro (2010). Heterogeneity, learning and information stickiness in inflation expectations. *Journal of Economic Behavior & Organization* 75(3), pp. 426–444. DOI: https://doi.org/10.1016/j.jebo.2010.05.012.

Preston, Bruce (2005). Learning about Monetary Policy Rules when Long-Horizon Expectations Matter. *International Journal of Central Banking* 1(2), pp. 81–126. DOI: 10.2139/ssrn.2482575.

Rotemberg, Julio J. (1982). Sticky prices in the United States. *Journal of Political Economy* 90(6), pp. 1187–1211. DOI: 10.1086/261117.

Schraeder, Stefanie (2015). Information Processing and Non-Bayesian Learning in Financial Markets. *Review of Finance* 20(2), pp. 823–853. DOI: 10.1093/rof/rfv002.

Slobodyan, Sergey and Raf Wouters (2012a). Learning in a medium-scale DSGE model with expectations based on small forecasting models. *American Economic Journal: Macroeconomics* 4(2), pp. 656–101. DOI: 10.1257/mac.4.2.65.

— (2012b). Learning in an estimated medium-scale DSGE model. *Journal of Economic Dynamics and Control* 36(1), pp. 26–46. DOI: 10.1016/j.jedc.2011.01.016.

Taylor, John B. (1993). Discretion versus policy rules in practice. In: *Carnegie-Rochester conference series on public policy*. Vol. 39, pp. 195–214. DOI: 10.1016/0167-2231(93)90009-L.

Wong, Arlene (2019). Refinancing and The Transmission of Monetary Policy toConsumption. *Working Paper*.

Yaari, Menahem E. (1965). Uncertain lifetime, life insurance, and the theory of the consumer. *Review of Economic Studies* 32(2), pp. 137–150. DOI: 10.2307/2296058.

Appendix

A. Equilibrium under EBL

We define equilibrium conditions for the EBL economy. Importantly, we perform aggregation of the cohorts' Euler equation into an aggregate IS-curve.

Aggregation. We follow the literature on heterogeneous expectations that relies on the axiomatic approach of Branch and McGough (2009) to be able to properly aggregate the decisions of agents with heterogeneous expectations without including the wealth distribution as an additional state variable.[31] To do so, we rely on two key assumptions from Branch and McGough (2009):

1. The structure of higher order beliefs: $\tilde{E}_t^i \tilde{E}_t^k x_{t+1} = \tilde{E}_t^i x_{t+1},\ i \neq k$.
2. Agents expect to return to the same wealth in the long-run: $\tilde{E}_t^i \left(\hat{c}_\infty - \hat{c}_\infty^i\right) = 0$.

Consider again the Euler equation given in (4):

$$c_{t|k} = \beta \tilde{E}_t^k \left(c_{t+1|k} \frac{r_t}{\pi_{t+1}} \right) .$$

The linearised Euler equation from a household in cohort i is given by:

$$\hat{c}_{t|i} = \tilde{E}_t^i \hat{c}_{t+1|i} - \left(\hat{r}_t - \tilde{E}_t^i \hat{\pi}_{t+1} \right) \ \forall i .$$

Forward iteration of the Euler equation yields:

$$\hat{c}_{t|i} = \underbrace{\lim_{j \to \infty} \tilde{E}_t^i \hat{c}_{\infty|i}}_{\equiv \tilde{E}_t^i \hat{c}_\infty^i} - \tilde{E}_t^i \sum_{j=0}^{\infty} \left(\hat{r}_{t+j} - \hat{\pi}_{t+1+j} \right) \ \forall i , \tag{20}$$

where we used Assumption A5. of Branch and McGough (2009) which states that the Law of Iterated Expectations is satisfied, i.e. $\tilde{E}_t^k \left(\tilde{E}_{t+1}^k \left(c_{t+2} \right) \right) = \tilde{E}_t^k (c_{t+2})$.
The aggregated linearised resource constraint in t and in $t+1$:

$$\begin{aligned} \hat{c}_t &= (1-\omega) \sum_{k=-\infty}^{t} \omega^{t-k} \hat{c}_{t|k} = \hat{y}_t \\ \hat{c}_{t+1} &= (1-\omega) \sum_{k=-\infty}^{t+1} \omega^{t+1-k} \hat{c}_{t+1|k} = \hat{y}_{t+1} . \end{aligned} \tag{21}$$

[31] Examples in the literature that rely in the axiomatic approach of Branch and McGough (2009) are for example Gasteiger (2014), Di Bartolomeo et al. (2016), Hagenhoff (2018).

Next, insert the forward iterated Euler equation (20) into the $t+1$-resource (21) for $\hat{c}_{t+1|k}$ (for each cohort in $t+1$, respectively) and take expectations of cohort i:

$$\tilde{E}_t^i \left[(1-\omega) \sum_{k=-\infty}^{t+1} \omega^{t+1-k} \left(\tilde{E}_t^k \hat{c}_\infty^k - \tilde{E}_t^k \sum_{j=1}^{\infty} (\hat{r}_{t+j} - \hat{\pi}_{t+1+j}) \right) \right] = \tilde{E}_t^i (\hat{y}_{t+1}) \ . \tag{22}$$

Following Branch and McGough (2009) and the exposition in Hagenhoff (2018) we note that how one thinks of higher-order beliefs matters for the further steps to aggregation. The former impose that agents' expectations about what other agents expect, are equal to their own expectation, which corresponds to assumption 1. Having departed from RE and, therefore, having assumed that agents do not know the underlying structure of the economy, imposing that they can not foresee how others form expectations is a natural step. Under assumption 1, we can rewrite (22) as:

$$\begin{aligned}
\tilde{E}_t^i (\hat{y}_{t+1}) &= (1-\omega) \sum_{k=-\infty}^{t+1} \omega^{t+1-k} \left(\tilde{E}_t^i c_\infty^k - \tilde{E}_t^i \sum_{j=1}^{\infty} (\hat{r}_{t+j} - \hat{\pi}_{t+1+j}) \right) \\
&= \tilde{E}_t^i \left((1-\omega) \sum_{k=-\infty}^{t+1} \omega^{t+1-k} c_\infty^k \right) \\
&\quad - \tilde{E}_t^i \left((1-\omega) \sum_{k=-\infty}^{t+1} \omega^{t+1-k} \sum_{j=1}^{\infty} (\hat{r}_{t+j} - \hat{\pi}_{t+1+j}) \right) \\
&= \tilde{E}_t^i c_\infty - \tilde{E}_t^i \sum_{j=1}^{\infty} (\hat{r}_{t+j} - \hat{\pi}_{t+1+j}) \ ,
\end{aligned}$$

where the last equality uses assumption 2, which we discuss below, and that weights sum to one. We use this to substitute the infinite sum of real interest rates in (20).

$$\begin{aligned}
\hat{c}_{t|i} &= \tilde{E}_t^i c_\infty^i - \tilde{E}_t^i \sum_{j=1}^{\infty} (\hat{r}_{t+j} - \hat{\pi}_{t+1+j}) - \left(\hat{r}_t - \tilde{E}_t^i \hat{\pi}_{t+1} \right) \\
&= \tilde{E}_t^i (\hat{y}_{t+1}) - \tilde{E}_t^i \left(c_\infty - c_\infty^i \right) - \left(\hat{r}_t - \tilde{E}_t^i \hat{\pi}_{t+1} \right) \ .
\end{aligned}$$

Of particular interest is the term $\tilde{E}_t^i (c_\infty - c_\infty^i)$, which denotes expected differences of own consumption and aggregate household consumption in the limit. Branch and McGough (2009) deal with such a term by assuming that agents agree on expected differences in limiting consumption so that in aggregation it vanishes. Equivalently, Hagenhoff (2018) assumes that agents expect to be back at the steady state in the long-run which also eliminates the term. We use the same assumption but adopt it for our usage in the following sense: take cohorts, i and k that both expect to have steady state consumption

in the long-run:

$$\tilde{E}_t^j \hat{c}_\infty^j = \tilde{E}_t^j \hat{c}_\infty \qquad \forall\, j = i, k\ .$$

Now, let cohort i take expectations of the limiting expectations for cohort k and invoke assumption 1: $\tilde{E}_t^i \left(\tilde{E}_t^k \hat{c}_\infty^k\right) \overset{2}{=} \tilde{E}_t^i \left(\tilde{E}_t^k \hat{c}_\infty\right) \overset{1}{=} \tilde{E}_t^i \hat{c}_\infty$. Cohorts not only expect to be back at the steady state but also expect this for others. Assuming agents expect to be back at the steady state in the long run, in which all consume equally, and having a unit mass of agents implies $c = c^k\ \forall\, k$ for non-explosive PLM-parameters. Under assumption 2:

$$\hat{c}_{t|i} = \tilde{E}_t^i \left(\hat{y}_{t+1}\right) - \left(\hat{r}_t - \tilde{E}_t^i \hat{\pi}_{t+1}\right)\ . \tag{23}$$

Note that (23) holds for all cohorts. Insert into the aggregate resource constraint in t:

$$\begin{aligned}\hat{y}_t &= (1-\omega) \sum_{k=-\infty}^{t} \omega^{t-k} \hat{c}_{t|k} = (1-\omega) \sum_{k=-\infty}^{t} \omega^{t-k} \left(\tilde{E}_t^k \left(\hat{y}_{t+1}\right) - \left(\hat{r}_t - \tilde{E}_t^k \hat{\pi}_{t+1}\right)\right) \\ &= (1-\omega) \sum_{k=-\infty}^{t} \omega^{t-k} \tilde{E}_t^k \left(\hat{y}_{t+1}\right) - (1-\omega) \sum_{k=-\infty}^{t} \omega^{t-k} \left(\hat{r}_t - \tilde{E}_t^k \hat{\pi}_{t+1}\right)\ .\end{aligned}$$

Finally, we use the definition of aggregate expectations, $\bar{E}_t x_{t+1} = (1-\omega) \sum_{k=-\infty}^{t} \omega^{t-k} \tilde{E}_t^k x_{t+1}$ to receive the aggregate dynamic IS-curve:

$$\hat{y}_t = \bar{E}_t \left(\hat{y}_{t+1}\right) - \left(\hat{r}_t - \bar{E}_t \hat{\pi}_{t+1}\right)\ . \tag{24}$$

It remains to rewrite the IS curve in terms of the output gap. We get

$$\tilde{y}_t = \bar{E}_t \left(\tilde{y}_{t+1}\right) - \left(\hat{r}_t - \bar{E}_t \hat{\pi}_{t+1} - r_t^n\right)$$

where $r_t^n = \bar{E}_t \left(\Delta \hat{x}_{t+1}\right) = 0$. The derivation of the NK Phillips Curves stays unchanged.

Summary. To sum up, we receive the following system of equations:

$$\tilde{y}_t = \bar{E}_t \tilde{y}_{t+1} - \left(\hat{r}_t - \bar{E}_t \hat{\pi}_{t+1}\right)\ , \tag{25}$$

$$\hat{\pi}_t = \beta\omega \bar{E}_t \hat{\pi}_{t+1} + \kappa \tilde{y}_t + u_t\ , \tag{26}$$

$$\hat{r}_t = \varphi_\pi \hat{\pi}_t + \varphi_y \tilde{y}_t + \epsilon_t^m\ , \tag{27}$$

$$\hat{\epsilon}_t^m = \rho \hat{\epsilon}_{t-1}^m + \nu_t^m\ , \tag{28}$$

$$u_t = \rho_u u_{t-1} + \nu_t^u\ , \tag{29}$$

where the expectations in the NKPC and in the dynamic IS-curve follow EBL.

B. Euler Equation Learning

To derive the equilibrium conditions under EBL, we rely on the widely used assumption in the SL-literature that agents choose consumption plans that satisfy the associated Euler equations. Under this so called *Euler equation learning*, agents base their decisions on the time t trade-off between current consumption and next period's consumption given the sequential budget constraint and on the one-period ahead forecast of consumption. Intuitively, each period agents equate the marginal costs of postponing consumption with the benefits, taking into account their current budget constraint and their one period ahead forecast of consumption and inflation.

Evans et al. (2013) argue that this formulation of optimisation behaviour is plausible. First, a forecast of own consumption in $t+1$ is required to make the period t consumption choice. It is helpful to consider what happens under REs. In a model with REs, the endogenous variables are a function of the relevant state variables in the economy. In our model, however, agents do not know the REE mapping and are unaware of the consumption function specified under REs. Instead, agents try to learn the RE mapping. Since the time t trade-off requires agents to form expectations about tomorrow's consumption plan without the knowledge of the RE mapping, the agent internalises how tomorrows consumption evolves given her PLM in mind. Given this PLM, the best "decision" of consumption tomorrow is plausibly a linear function of current consumption.

Further, agents are assumed to think just one period ahead without explicitly taking into account the inter-temporal budget constraint. Yet, Evans et al. (2013) show that for a convergent learning process, the inter-temporal budget is satisfied. Since along the sequence of *temporary* equilibria, agents consumption is equal to its income and by imposing a solvency condition like the one in equation (3), ex post consistency in the accounting over the planing horizon of the household that has some strictly positive probability to live forever is fulfilled. For further details see Evans and Honkapohja (2012).

Convergence of the learning process is an important property of the SL-literature. Usually, agents learn based on a correctly-specified PLM so that one can derive the conditions for convergence to the RE equilibrium (called *E-stability*). Yet, agents in our model learn based on a misspecified model of the economy, so that convergence to the REE is not possible. Instead, Evans and Honkapohja (2012) and Berardi (2009) discuss the concept of a restricted perceptions equilibrium and of a heterogeneous expectations equilibrium, respectively. Both concepts describe the situation in which agents' beliefs converged to a fixed point that, while not being the REE, allows to make the best possible forecast given the (misspecified) model they use. In our model, expectations are heterogeneous and based on a wrong PLM. Yet, through simulation we show that PLM parameters converge to a non-explosive value for old agents (see Figure 5) so that expectations are stable.

C. Algorithm

As described in section 3, the vector of all endogenous variables, Y, contains a subset variables that appear with lead, Y^f. Households in cohort k forecast future values of variables $\mathbf{z} \in Y^f$ using the following linear regression model:

$$\mathbf{z}_{t|k} = b^{\mathbf{z}}_{t-1|k}\mathbf{z}_{t-1} + \varepsilon_{t|k} \ , \tag{C.1}$$

where $b^{\mathbf{z}}_{t-1|k}$ is estimated via recursive least squares according to:

$$b^{\mathbf{z}}_{t|k} = b^{\mathbf{z}}_{t-1|k} + \gamma_{t|k}\left(R^{\mathbf{z}}_{t|k}\right)^{-1}\mathbf{z}_{t-1}\hat{\varepsilon}^{\mathbf{z}}_{t|k} \tag{18a}$$

$$R^{\mathbf{z}}_{t|k} = R^{\mathbf{z}}_{t-1|k} + \gamma_{t|k}(\mathbf{z}_{t-1}\mathbf{z}'_{t-1} - R^{\mathbf{z}}_{t-1|k}) \ . \tag{18b}$$

where $\hat{\varepsilon}^{\mathbf{z}}_{t|k} = \mathbf{z}_t - b^{\mathbf{z}}_{t-1|k}\mathbf{z}_{t-1}$ denotes the forecast error of an individual in cohort k for the variable $\mathbf{z}$.

Algorithm

1. Simulate the exogenous process w_t.[32]
2. Simulate the economy under RE for an initial phase of $T_{\text{init}} = 120$ periods to get initial values $b^{\mathbf{z}}_{-1}$ and $R^{\mathbf{z}}_{-1}$.[33] In period $T_b + 1$, we endow all cohorts with the same $b^{\mathbf{z}}_{-1|k} = b^{\mathbf{z}}_{-1}$ and $R^{\mathbf{z}}_{-1|k} = R^{\mathbf{z}}_{-1}$ and switch to EBL.
3. We insert the PLM parameters into `dynare` and obtain policy functions.
4. Using the policy functions, we simulate a new observation of the endogenous variables in Y using w_t from 1.
5. Based on the new observation of $\mathbf{z} \in Y^f \subset Y$ we update $b^{\mathbf{z}}$ and $R^{\mathbf{z}}$ for each cohort using (18a) and (18b).
6. We repeat steps 3.-5. for $T_{\text{sim}} = 10,000$ periods.

The actual simulation displayed in the graphs discards the first iterations under EBL to allow the impact of initial values from the RE economy to die out. When we use `dynare`, we need to specify the number of cohorts in the mod-file. While, theoretically, agents can live for infinity (and number of cohorts is infinite), we need to choose a finite number when specifying the model. A high number approximates the "true" economy closer but

[32] For the simulation we rely on Matlab and `dynare 4.6.2`.

[33] They follow from a simple least squares regression of $\mathbf{z}$ on its lagged value. In that sense initial beliefs are close to those implied by the RE model.

comes at the cost of computational complexity. We, therefore, choose the number of cohorts as 200 (equivalent to a 50-year working life) and normalise cohort weights to sum to one. We endow the new born cohort with PLM parameters equal to the aggregate PLM parameter of the last period. A variation of the initial belief did not greatly alter results (see Appendix D).

In step 6., we use a so-called projection facility (PF, henceforth) to ensure the model with EBL can be solved. Conceptually, it reinitializes the updating step as soon as new simulated data implies unstable PLM parameters. We proceed in 2 steps:

Projection Facility

1. We take the updated PLM coefficients and check whether they make they forecasting model explosive, $|b^{\mathbf{z}}_{t|k}| > 1$. If the forecasting model generates non-explosive behaviour, we allow the updating step.

2. If behaviour is explosive we proceed as follows:

   (a) The problem may lie with the "effective gain" $\gamma_{t|k}(R^{\mathbf{z}}_{t|k})^{-1}$: updating in (18a) may also for a small gain $\gamma_{t|k}$ lead to explosive values if the matrix $R^{\mathbf{z}}$ has its smallest eigenvalue close to zero. As outlined by Slobodyan and Wouters (2012a) this might occur when initial beliefs are derived from the REE. We follow the authors and invoke a "Ridge correction" for those cases: if the smallest eigenvalue of $R^{\mathbf{z}}$ is smaller than a constant ι, the inverse of the moment matrix $(R^{\mathbf{z}})^{-1}$ is replaced by $(R^{\mathbf{z}} + \iota\mathbf{I})^{-1}$, where $\iota = 10^{-5}$.

   (b) If either the smallest eigenvalue of $R^{\mathbf{z}}$ did not invoke above correction or if despite correction a household's forecasting model is explosive, we invoke the "standard" form of the PF and ignore the updating step.

Two comments are at order. First, our usage of the PF deviates from the "standard" procedure that ignores the updating step if behaviour is explosive. Any time the PF is used, information and agent behaviour (generated by our model) are discarded. Using the correction for the effective gain before the "standard" PF, constitutes less of an intervention into the model. Furthermore, although our gain parameter is decreasing, young agents, by construction, have high gain parameters. Thus, in case the forecast error turns out to be extremely high (e.g. due to a high shock), especially estimates of young agents are easily drifting to extreme PLM parameters. While we desire strong updating behaviour of young agents, we face the trade-off against having extreme PLM parameters. Therefore, we still invoke the PF that ignores updating steps that lead to parameters ± 1. Second, using the PF can be rationalized by (realistically) assuming that forecasters do not use explosive models.

Again, to compute the initial parameters for agents' PLM, we simulate the model with RE for $T_{\text{init}} = 120$ periods and obtain $\delta^{\mathbf{z}}_{-1}$ by estimating an AR(1) process for $\mathbf{z} \in \{\tilde{y}, \hat{\pi}\}$. In the next step we simulate both the model with EBL under a supply shock for $T_{\text{sim}} = T_b + T_{\text{irf}}$ periods as described in 3–6, while setting the monetary policy shock to zero. We then simulate the model under the *same* path of supply shocks and add an innovation of 25 basis points to the monetary policy shock at time $T_{\text{imp}} = T_b + 1$. We then take the difference between these two series as the impulse response function to a monetary policy shock. We repeat this exercise $8,000$ times and take the mean response as the final impulse response function to a monetary policy shock. In an analogous we way, we compute the impulse response function to a MP shock in the model with CGL.

D. Sensitivity Analysis of the Perceived Persistence

In this appendix we check our results regarding the aggregate perceived persistence for different parametrisations of the model and a variation of initial beliefs.

Robustness w.r.t. θ. The gain to new information is age-dependent under EBL. To parametrise its behaviour across age groups we use the estimated shape parameter θ of Malmendier and Nagel (2016). While we set $\theta = 3.044$ to the estimate of their preferred specification, the authors also provide values for different regression specifications. In fact our choice of θ is the lowest value provided so that we check the robustness of our results against the highest estimate in Malmendier and Nagel (2016), $\theta = 4.144$.
Recall that the shape parameter governs how quickly gains reduce across age groups (see Figure 3). A higher value, hence, will intensify the difference between young and old agents, which is the root of expectation differences and of our results. Figure 12 compares the distribution of aggregate PLM parameters under both specifications of θ. As expected, we observe dispersion of parameters to increase as the gain is set to the highest value estimated by Malmendier and Nagel (2016). In that sense, our results constitute a lower bound. In particular, the difference to a model with CGL will increase as results under CGL are unaffected by our choice of θ.

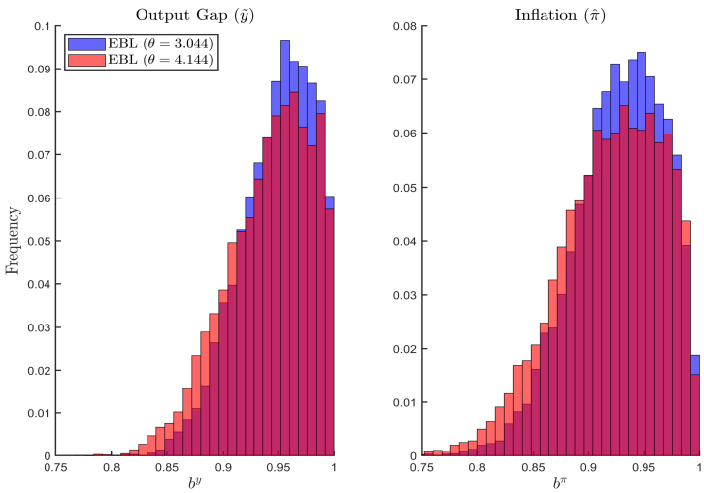

FIGURE 12: Distribution of PLM Parameters: Variation of θ

Notes: The blue histogram denotes the distribution of PLM parameters in the model under baseline calibration ($\theta = 3.044$) while the red histogram shows the distribution of the PLM parameters for $\theta = 4.144$. The left part shows the PLM parameter for the output gap while the right part shows the PLM parameter for inflation.

Robustness w.r.t. κ. In the next step, we explore the sensitivity of the results with respect to the slope of the NKPC, κ. Such a variation can result from a change in:

1. the elasticity of substitution of intermediate goods, ε,

2. the Rotemberg parameter ϕ, and

3. the DRS parameter in the production function, α.

We investigate the effect of a decrease in κ by increasing α to one. In this case, inflation becomes less sensitive to variations in output. As can be seen in Figure 13, the mean of the distribution of the PLM parameter for both inflation and output in the model with EBL is lower relative to the corresponding value in the model with CGL. Hence, the key mechanism that the perceived persistence of both inflation and output in the model with EBL is lower, on average, relative to the model with CGL is unaffected. We verified that, under EBL, the stabilisation trade-off of monetary policy remains less severe relative to the model with CGL as κ decreases. Moreover, also the transmission of monetary policy shocks under EBL remains both less pronounced and less persistent relative to the one in the model with CGL.[34]

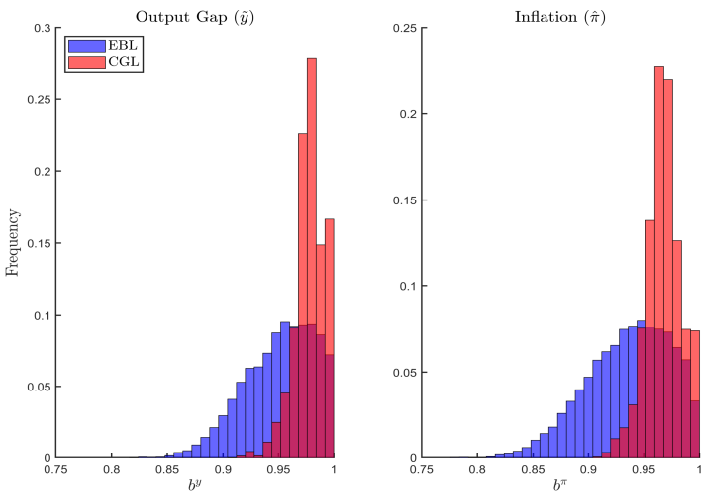

FIGURE 13: Distribution of PLM Parameters: Variation of κ

Notes: The blue histogram denotes the distribution of PLM parameters in the model under EBL if $\alpha = 1$ while the red histogram shows the distribution of the PLM parameters in the model with CGL for $\alpha = 1$. The left part shows the PLM parameter for the output gap while the right part shows the PLM parameter for inflation.

Robustness w.r.t. the initial belief. Each new cohort needs to be given an initial set of PLM parameters to start the RLS updating process. In the main text we assume that each new cohort uses previous period's aggregate PLM parameter. Yet, initial beliefs may be of relevance for the general result with respect to the aggregate persistence as they change the young agents' starting point in updating parameters.

[34]The corresponding simulation results are available upon request.

Now, we instead assume that each new born cohort draws its initial parameter from a normal distribution around the RE estimate, where the normal distribution is truncated at ± 1. Figure 14 shows the histogram of aggregate PLM parameters of inflation and the output gap for the models with EBL and CGL. The key result that the aggregate persistence is smaller and more dispersed under EBL remains unchanged.

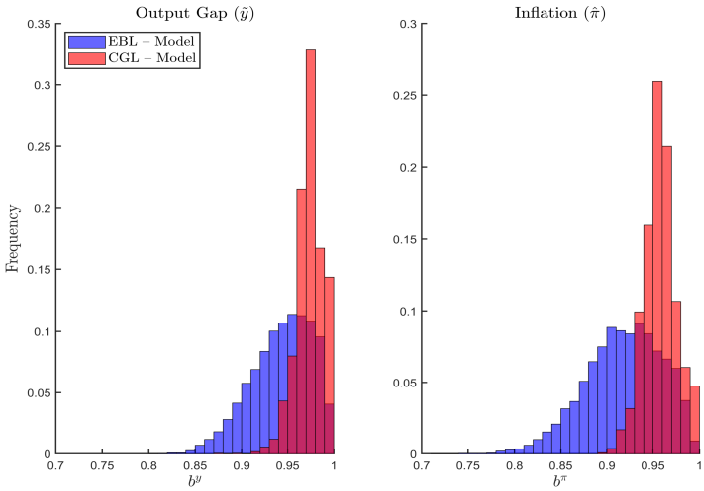

FIGURE 14: Distribution of PLM Parameters: Variation of Initial Beliefs

Notes: The blue histogram denotes the distribution of PLM parameters in the model under EBL while the red histogram shows the distribution of the PLM parameters in the model with CGL. The left part shows the PLM parameter for the output gap while the right part shows the PLM parameter for inflation.