I would like to thank Rodrigo Guimarães and Lennart Brandt for assistance in preparing this speech, and for comments and discussions I would like to thank Andrew Bailey, Torsten Bell, Ben Broadbent, Maren Froemel, Richard Harrison, Andrew Hauser, Sophie Piton, Michael Saunders, Fergal Shortall, Silvana Tenreyro, and Lukas von dem Berge.

All speeches are available online at www.bankofengland.co.uk/news/speeches and @BoE_PressOffice
1. Introduction

Good afternoon. Today’s public lecture will be my last as an external member of the Monetary Policy Committee, since my term comes to an end shortly. It has been a privilege to have served on the MPC in the past six years. When I joined the Committee, someone made the comment that it might not be that exciting a period in history to do this job, because the expectation at the time was that the economy would just continue to recover gradually from the financial crisis, and that interest rates would only rise very slowly, over a number of years. As it turns out, we had Brexit, a US-China trade war, and a global pandemic. Policy rates have gone down, up, and down again, to levels that were not even thought to be on the list of policy options back in 2015. QE has re-started, stopped, and re-started again. It has been an action-packed six year term!

I want to extend my thanks to my colleagues on the MPC, and to the Bank of England’s wonderful staff.

These are some of the smartest and most dedicated people I have ever met, and it has truly been a pleasure to work with them.

In my first speech as an MPC member, I talked about why we had ended up in a low interest rate environment, and whether it would last. I argued the following:

Structural developments in indebtedness, demographics and the distribution of the income (3D) had resulted in an environment where low interest rates would prevail for years, possibly even decades, even with growth at potential and inflation at target.

I suggested two implications. One, that we should factor this thinking into our forecasting models, which otherwise would continue to predict rapid growth in activity and prices, which could lead to costly policy mistakes. Two, that when the time comes for tightening policy, we should proceed cautiously, as policy space was asymmetric and the distance to neutral was small. The international experience had shown it was difficult to correct for premature tightening.

Today I want to revisit that topic.

I will organise my thoughts along two themes.

First, what have we learned since then, both in terms of economic theory and empirical evidence, about these structural developments?

Second, what are the policy implications? As I will shortly no longer be an MPC member, I’ll allow myself the freedom to discuss policy implications that go beyond the strict remit of the MPC.
2. What have we learned: demographics

The key underlying developments in demographics are the fall in the fertility rate and the rise in longevity: fewer babies and longer lifespans means the population is, on average, getting older. That has been understood for a long time, but a widespread discussion of the macroeconomic implications is far more recent. My focus is on the implication for interest rates. In particular, the implication for the neutral interest rate, or $r^*$, by which I mean the interest rate that is expected to prevail when output is sustainably at potential and inflation is at target.

The main insight is that this demographic transition is primarily about desired stocks of assets of the whole population, not about the flow of savings, or savings rate, of the old. Once we understand that, the sign of the effect of demographics on interest rates is clear. The intuition of many observers (including me, some years ago) is that this is a story about old people retiring, about baby-boomers. Just before retiring, old people save a lot, so push down on $r^*$. Once they cross the retirement threshold, they become dis-savers, so push up on $r^*$. But this is a very partial analysis. What dominates quantitatively is not the fact that more old people will soon be crossing the retirement threshold, but that a growing share of the population has a higher desired stock of assets to finance their retirement. That will continue for many decades and will, if anything, add further downward pressure on $r^*$ relative to today.

Chart 1: Average wealth by age group (GBP thousands)

Sources: UK Wealth and Asset Survey and Bank calculations. Notes: Chart shows weighted average wealth by age group from round 6 of the UK Wealth and Asset Survey. This data was collected between 2016 and 2018.

1 See Auclert et al (2021) for a recent and comprehensive review of the literature on the macroeconomic effects of demographics, which focused on aggregate savings, pensions, asset prices and capital flows, among others. Auclert et al (2021) show the implications of demographic for changes in interest rates, global imbalances and aggregate wealth accumulation in a unified framework in theory and the implications for a large panel of countries.
2 The higher demand for assets coming from the larger desired stock of assets of older people is substantially larger than the lower demand coming from the flow effect of saving out of income after retirement.
3 For a rich discussion of demographics and savings that comes to a different conclusion, see Goodhart and Pradhan (2020).
Chart 1 shows the average asset holdings of individuals, by age. We can see that people accumulate assets as they age, and after retirement they run down assets. This hump-shaped pattern is common across many countries and over time. Crucially, retirees do not run down their asset holdings very quickly. In contrast, people who enter their 50s, for example, hold far more assets than when they were in their 40s. What I will demonstrate is that the latter effect totally dominates the former effect: the additional saving of the middle-aged outweighs the modest dissaving of the retirees.

We can combine data on life-cycle asset holdings with data on the future age distribution of the population, to project forward the level of asset holdings.

Chart 2 shows how the UK age distribution is changing over time. I am showing three snapshots: 1980, 2020, and 2060. We can see that, between 1980 (grey line) and now (blue line), the biggest change in the age distribution has been a rise in the size of the 30-60 age group. We can also see that, in the next four decades, the biggest projected rise is in the size of the 60-90 age group.

**Chart 2: UK age distribution over time (thousands of population)**

![Age distribution chart]

Source: UN Population Prospects.

Putting it all together, we can project today’s total assets backwards and forwards using the change in the age structure of the population. Specifically, we can ask the question: what do asset holdings look like at some future (or past) date, if the life cycle profile of individual asset holdings remains constant? This analysis therefore shows us the compositional effect on asset holdings from changes in the age distribution (see Lisack et al (2021) and Auclert et al (2021)).

---


5 This empirical analysis focuses on the pure composition effect, keeping the life-cycle wealth profile constant. There is an additional effect, reinforcing the dynamic, from those of a given age deciding to hold more assets in response to rising longevity.
Chart 3 shows the result. We can see that, starting around 30 years ago, the ageing effect started pushing up more rapidly on per capita asset holdings. We can also see that this process is by no means over, and is not projected to reverse. In fact, we are only about two thirds of the way through this demographic savings transition. The impact on r* depends somewhat on the extent to which economic agents are forward-looking and anticipate this change. But either there is further downward demographic pressure on r* to come, or r* simply remains low. There is no upward pressure from demographics, soon or in the distant future.

Chart 3: Stock of total wealth per capita at current wealth-age profile (index 2020 = 100)

Sources: UN Population Prospects, UK Wealth and Asset Survey, and Bank calculations. Notes: Total stock of wealth is calculated by holding the current wealth-age profile (from round 6 of the WAS) constant and multiplying with the respective age profile of the latest UN Population Prospects. To control for overall population growth, the aggregate is then divided by total population.

I am showing this calculation for the UK, but a number of research papers have made similar calculations in the past few years, for specific advanced economies or for aggregates of advanced economies. All lead to the conclusion that the transition is far from over.

What is still missing from these models is a distinction between risky and safe assets. There is a good reason for this: adding a complex portfolio choice to already rich and complex models is really difficult. I certainly do not know how to do it. But a further insight we might gain is to understand the additional effect from the fact that older people hold not only more assets, but more safe assets relative to risky assets. This could explain why we have not observed a uniform fall in all rates of return, as the simple model predicts, but instead we have observed a fall in risk-free rates of return, while risk premia have been persistent.

---

7 As I discussed in Vlieghe (2016a), the demographic transition in Japan started nearly two decades earlier, and Japanese yields fell then (even as global yields remained around 5% for another 15 years) just as the yields in countries going through similar demographic transitions have experienced now. This suggests the local demographic factors are key, and also supports the theoretical prediction that the downward pressure on interest rates is far from temporary.
8 See Yogo (2016) for a model calibrated to US data jointly explaining asset allocation (the large drops in risky asset holdings - equity and housing - and increase in safe fixed-income asset holdings) and health expenditures in the US in retirement; and Fagereng et al (2017) documenting similar patterns using administrative records for Norway, with an additional focus on stock market participation (suggesting households rebalance their portfolio away from stocks before retirement and exit the stock market after retirement).
3. Income inequality, debt and risk

Next, I want to turn to the macroeconomic effect of rising income inequality, as measured for example by the income Gini coefficient or the share of income earned by the top 10% or 1% of earners. Rising income inequality has been experienced by most advanced economies over the past four decades and more recently also by many large emerging economies (Dabla-Norris et al (2015)). Let us first recall briefly the intuition for the macroeconomic effect of changes income inequality, which is beautifully simple, yet missing from most standard macro models. Those earning a higher income have a lower marginal propensity to consume. A rise in inequality means more income earned by those who have a lower marginal propensity to consume. This therefore requires a lower interest rate to maintain aggregate consumption demand.

Several papers have been published that integrate this mechanism into a macro model to gauge its quantitative importance. One relatively recent additional feature in the literature is that higher income inequality can ultimately lead to higher wealth inequality, which has a bigger effect on r*. This is another case in which the stock effect (in this case the stock of savings) can be sizeable and dominant, but it is often less of a focus in standard models because it is technically complex to implement. Like demographics, the changes in wealth distribution are also slow-moving, which makes it empirically difficult to unpick their effect.

A second insight is the interaction between income inequality, debt, and risk, three factors that I have previously discussed separately, but which have been shown to interact with each other and reinforce each other in powerful ways.

In my first speech (Vlieghe (2016a)) I discussed indebtedness from the perspective of a debt overhang that, after a downward revision to future income prospects, leads to a period of deleveraging, associated with weakened demand. I used to think of this as a more temporary force than demographics or inequality, related specifically to the credit boom and credit bust in the years before and after the global financial crisis.

Now several papers analyse the interaction between debt and the distribution of income, something which I think Kumhof et al (2015) were first to formalise. The mechanism is as follows. Those at the top of the income distribution not only have a lower propensity to consume, but they also desire to accumulate more wealth. The higher desired asset holdings push down on r* not just via a higher capital to income ratio, but also via an increase in lending to lower income households. The interest rate falls far enough to encourage lower income households to borrow. In effect, higher income inequality looks like a positive credit supply

---

9 It implies an additional “state variable” to keep track in equilibrium, which makes solving the models substantially harder. For this reason usually assumptions are made such that stocks (or ratios relative output) don’t change in equilibrium and all that matters are the flows around the equilibrium ratios.

10 See Mian et al (2021), Rannenberg (2020), and Cairo and Sim (2018).

11 This is accomplished by either having wealth in the utility function (as in Cairo and Sim (2018) or Rannenberg (2020), among others), so agents desire to accumulate more wealth, or non-homothetic consumption behaviour (as in Mian et al (2021)), where agents have decreasing marginal propensity to consume out of permanent income (as opposed to the usual assumption of decreasing MPC out of temporary income).
shock for lower income households. The corresponding assets representing this credit to lower income households are held, directly or indirectly, by the higher income households.

In this mechanism, debt is not an independent factor, but rather the endogenous consequence of income inequality. There are multiple channels through which income inequality, via higher debt, reduces \( r^* \).

First is the credit supply effect from the higher demand for assets (or, conversely, supply of credit) by the higher income households that I already mentioned.

The second effect is a risk effect. Higher debt, in this example higher household debt, makes the economy more fragile, more prone to a financial crisis, even if the debt itself increases welfare for the borrowers by easing a constraint on their investment or durables consumption. That results in a skewed and fat-tailed distribution of future macroeconomic outcomes, which leads to a lower risk-free rate while maintaining a high risk premium.\(^{12}\)

Third is what Mian \textit{et al} (2021) call an “indebted demand” effect. Having accumulated debt to fund consumption today, lower income households will need to reduce future demand in order to be able to service the higher debt.

And once we combine these ideas with a lower bound on interest rates, and therefore an asymmetry in the ability of monetary policy to respond to swings in the business cycle, the risk effect can become larger.

Higher debt leads to a higher probability and severity of a crisis and therefore a more skewed and fat-tailed distribution of macroeconomic outcomes. The steady-state level of \( r^* \) will therefore be lower, which reduces the space for monetary policy to respond to a crisis. In turn, that means the crisis outcome is likely to be worse, so the distribution of macroeconomic outcomes becomes even more skewed, leading to an even lower level of \( r^* \), and even less monetary policy space to respond to a future crisis.\(^{13}\)

Even in an environment without a risk channel, Mian \textit{et al} (2021) suggest another possible low interest rate trap. This is somewhat more speculative, and I think it needs further work to check how general the conclusion is, but I mention it as a possibility. If, after a recession or crisis, monetary policy is sufficiently stimulative to push up debt, this will lower future \( r^* \) via the indebted demand effect. Low interest rates beget even lower interest rates in the future, in this view of the world.


\(^{13}\) A number of studies have shown importance of ELB in magnifying the macroeconomic effects of debt deleveraging (see Benigno \textit{et al} (2020)) and changes in income distribution (see Cairo and Sim (2018) and Auclert and Rognlie (2020)).
We have not experienced a monetary-policy-driven increase in debt recently, I think. But a look at private sector credit data in Chart 4 does reveal that, although debt to income fell significantly in the first five years after the financial crisis, it subsequently stabilised near historically high levels. The US saw similar developments. There was no further deleveraging after 2015. Monetary policy was trying to support demand enough to keep inflation anchored, while macroprudential policy was trying to prevent the kind of rise in leverage that would endanger financial stability. One could think of the joint objective of these policies as threading a fine line between stimulating demand enough to prevent a deleveraging slump, but not so much that an increase in leverage actually starts threatening financial stability.

**Chart 4: Private and public debt in UK and US (% of nominal GDP)**

Even merely preventing debt from falling, rather than pushing it up, can result in a reduction in future monetary policy space. This is because more shocks might come along that might push debt higher, leading to a ratchet effect over time.

It is important to emphasise that, even in this world, reducing interest rates to ensure demand is strong enough to hit the inflation target is the optimal monetary policy response, other things equal. That is, even knowing what we know now, and even if we fully subscribe to this hypothetical mechanism, easing monetary policy in response to demand shocks was and will continue to be the right policy. Cutting interest rates by less would not solve this structural problem. It would have made the downturn much worse. But it is possible that, by increasing debt, or preventing debt from falling, current monetary policy actions reduce future monetary policy space. The key is to remove the ‘other things equal’ constraint: rather than respond to demand shocks largely with monetary policy stimulus, another option is for fiscal policy to play a different

---

14 I am talking here about the decline and subsequent stabilisation of debt ratios since the financial crisis. The jury is still out on the post-pandemic debt developments. So far, private sector debt has risen, but by less than suggested by chart 4. Much though not all of the spike in in the debt ratio at the end of the sample period is due to the sharp but likely temporary fall in nominal GDP (data to Q4), the denominator of the debt ratio.

15 The view that lower policy rates in response to recessions have contributed to the trend decline in real interest rates has been articulated by Claudio Borio and co-authors at the BIS (see Borio and Distayat (2014)).
role, both in addressing structural factors that lead to inequality and indebtedness, and in dealing with recessions (automatic stabilisers). I will come back to this.

While I think the 3D drivers are the most important ones for understanding developments in $r^*$, there have been many other important macro trends that are of interest in their own right, and likely have some additional effect on $r^*$ as well. I have omitted them from my discussion merely for the sake of brevity. I am thinking about such developments as possible changes in the labour share of income, the fall in the relative price of capital, the increased importance of intangibles, increased industry concentration, increased automation, and the effect of all of these on productivity.

4. Policy implications: QE and limited policy space

Low neutral rates means limited space for cutting the policy rate when the economy needs stimulus. When policy rates were at or near their effective lower bound (ELB), the MPC and many other central banks relied on QE to add further stimulus.

But even the amount of stimulus provided by QE ultimately has a limit as well, which arises from two sources. First, even long-term yields have a lower bound. Second, the power of QE is highly state-dependent.

a. QE headroom and the level of long term yields

Whether you think QE works mainly via a persistent portfolio balance effect that reduces risk premia, or – as I do – mainly by lowering future rate expectations and raising inflation expectations back to target, with a powerful temporary liquidity effect in dysfunctional markets, in either case the persistent effect of QE works by lowering long term yields.

Once long term yields fall to very low levels, approaching the effective lower bound on policy rates, yields cannot fall much further. Such conditions prevailed in H2 2020 when UK 10 year yields averaged around 0.2%. Additional QE beyond that point does not, in my view, deliver significant additional stimulus, because it cannot lower yields significantly further.\(^\text{16}\)

It is worth spelling out this argument in detail. The most rigorous argument, which does not rely on any model of QE and risk premia, is a so-called no-arbitrage argument. If long-term yields were below the expected lower bound, then you could make a certain profit by borrowing at long-term rates (or shorting the long bond) and reinvesting the proceeds at the short-term rate. Normally such strategy is risky because short-term rates could fall, generating a loss. But if short-term rates can never fall below the lower bound, the strategy is

\(^\text{16}\)Additional QE might still be able to prevent a rise in yields where such a rise would represent an unwarranted tightening that risks keeping or pushing inflation below its target. I elaborate on this in the next section.
guaranteed to earn more than what you need to repay. Therefore, long-term yields cannot fall below the ELB, or there would be an arbitrage opportunity (a profitable strategy that can never lead to a loss).\footnote{The mechanism through which this arbitrage argument works is the following: since selling the long bond and investing in short bonds can only lead to a profit without any risk of loss if the long term interest is below the ELB, investors have an incentive to keep selling the long bond and buying shorter bonds until the price of the long bond falls enough such that the yield would no longer be below the ELB (and the arbitrage opportunity disappears). This does not rely on arbitrageurs actually selling the long bond. As long as investors realise they are strictly better off buying only short term bonds, demand for long term bonds would collapse, lowering long term bond prices (hence raising their yields).}

Note that this line of reasoning requires the lower bound to be known and fixed. The argument does not apply exactly if there is a possibility of lowering the lower bound in the future. But in that case what is driving bond yields lower is not the buying power of QE per se, but speculation about the possibility of lower policy rates in the future.

Is there not a possibility that term premia could be negative, i.e. that bond yields are lower than the expectations of future policy rates? This can indeed happen,\footnote{See Guimarães (2012, 2016) and Vlieghe (2018) explaining that this is in fact normal if inflation and consumption are positively correlated, which has been the case in the UK since early 2000s.} but only if expectations of future policy rates are above the lower bound. A negative risk premium reflects the insurance value of a long-term bond relative to the short-term interest rate. If expected future short-term rates can fall, then the long-term bond will gain more than a short-term bond, hence it will have a lower (and possibly negative) risk premium if those excess gains are negatively correlated with consumption (i.e. gains happen in bad states of the economy). But once expected future short-term rates are at the lower bound (so the expected rate component of the long term bond is at the lower bound), they cannot fall further, hence the long-term bond cannot act as insurance relative to short bonds, and risk premia cannot be negative anymore. Not surprisingly we reach the same conclusion as with the no-arbitrage argument.

b. State-dependent power of QE

As I have argued on several previous occasions,\footnote{See Vlieghe (2016b, 2018, 2020, 2021).} the main and persistent effect of QE has come through lower expected real rates, keeping inflation expectations anchored and lowering expected nominal yields by revealing our reaction function at the ELB. Beyond expectations, I believe QE has a liquidity channel (through the level of reserves in the banking system) and a temporary term premium effect, but this temporary term premium effect is much larger during periods of market turmoil than when financial markets are functioning smoothly. That is exactly what the theory predicts. During periods of orderly market functioning, arbitrageurs have sufficient balance sheet capacity to absorb buying and selling flows in financial markets, so that these flows have only a small and short-term effect on the prices of financial instruments.

But when arbitrageurs’ balance sheets are constrained, for example when they have suffered losses after a period of sharp declines in the prices of major asset classes, market liquidity is low, and the central bank can have a larger (term premium) impact on prices by buying bonds and increasing reserves (liquidity). That can result not just in a fall in government bond yields, but also in falls in risk spreads and falls in volatility in risky...
asset markets, generally acting to ease financial conditions. Without QE, the economic downturn in 2009 and 2020 would have been exacerbated by the tightening in financial conditions, reinforcing the shortfall in activity relative to potential, and ultimately the shortfall in inflation relative to our 2% target. QE helps to prevent the damage to the economy that would otherwise have taken place. But this particular QE channel does not add new stimulus relative to the pre-turmoil state of the economy.

Chart 5 shows, on the vertical axis, the movement in long-term bond yields on the day of major MPC QE announcements. On the horizontal axis I show the extent to which the QE announcement was a surprise.

The red line illustrates the quantitative impact of QE on market yields, based on the first QE programme in early 2009. The other dots are various subsequent QE announcements. I have also labelled the March 2020 QE announcement.

To me, this chart illustrates that the impact of QE on yields is not constant. There were large yield effects from the early 2009 and March 2020 QE announcements, in particular in early 2009, but much smaller effects from all the other programmes. In fact, if you exclude the early 2009 and March 2020 observations from the sample, the slope of the fitted line through the other QE announcements is close to zero.

---

20 The chart shows the 2 day change in 10 year (zero-coupon) yields for each announcement day against measures of QE surprises built using surveys of QE expectations. Using instead the average in the 2 day change for yields with maturities from 5 to 25 years gives very similar results (see Joyce et al (2011), Haldane et al (2016), Broadbent (2018), and Froemel et al (2021)).

21 The only reason to focus on surprises is that the extent of the QE surprise allows us to identify QE as causal factor for yields on that day. Changes in expectations of QE can affect yields all the time, but are much more difficult if not impossible to distinguish from non-policy effects on yields, such as changes in the economic outlook based on data news.

22 The ‘March 2009’ observation combines the reaction to the February and March 2009 announcements, assuming the combined effect of those announcements was to increase expectations to £142bn of QE (which is the first survey measure available, after the March announcement). Because there was discussion of QE in the UK from as early as November 2008 (following the Fed’s start of their LSAP program), this represents a conservative upper bound on the size of surprise in these two announcements, which makes the yield impact (size of yield changes divided by the surprise) on those days a conservative lower bound of the QE price impact in early 2009. The first LSAP announcements in the US also resulted in the largest yield responses (see Swanson (2021)).
Chart 6 illustrates what was special about the early 2009 and March 2020 episodes: they took place when market functioning was severely disrupted. The other QE announcements took place against a background of relatively smoother market functioning. This analysis supports the idea that the power of QE is state-contingent: QE had powerful effects when market functioning was poor, but had much less impact when market functioning was good. This analysis is obviously based on just a few data points (though also holds in the US), but it is exactly the pattern of state dependence that should be expected according to the model of Vayanos and Villa (2020), the theoretical basis for almost all analysis of QE effects on yields to date, where the impact of any bond supply or demand shock depends on the risk capacity of arbitrageurs (their capital and risk aversion).

c. Implications for monetary policy strategy

I want to draw out three important implications of this preceding discussion for how we should think about QE strategy.

First, QE headroom should be measured in basis points, not billions. While QE can help add some stimulus when the policy rate is at the lower bound, once long-term yields fall close to the lower bound as well, additional QE will not lower long-term yields further, so will not add further stimulus in well-functioning markets.

Second, pace matters. This is a simple general point that should not be controversial: if we buy 1% of the gilt market in 1 month we should expect a much larger effect than if we buy 1% over a 3 year period. So a faster pace should, other things equal, lead to larger effect on yields. In addition, because QE creates reserves, it provides the ultimate source of liquidity to the banking sector. Therefore, when markets are disrupted and aggregate liquidity is low, a fast purchasing pace of QE is particularly powerful. Once market functioning is restored, however, continued purchases at a fast pace are not necessary anymore, and slow purchases likely have little effect beyond their signalling value that future rate hikes are still some way off. This is why I pointed out in my last speech (Vlieghe (2021)) that I did not expect the additional QE beyond last year’s period of market disruption to add additional stimulus to the economy. It provided some insurance against a withdrawal of stimulus that might otherwise have taken place, either via unwarranted expectations of near-term rate hikes, or via a renewed period of market disruption. The future stock of QE matters mostly to

---

23 The chart shows the VIX, a measure of equity market volatility commonly used to gauge market stress, and the financial intermediary capital measure of He et al (2017), who have shown this is a good proxy for arbitrageur capital that affects the prices of all asset classes and contains information beyond other proxies for aggregate liquidity. For both measures, a higher value reflects less liquid, or distressed, markets.

24 By which I mean basis points of long-term yields, i.e. how far longer term yields are from the ELB, not billions of pounds of government bonds that remain available for central banks to buy.

25 I have previously highlighted (Vlieghe (2018, 2020)) that, on top of the effect on expected path of rates and temporary term premium effects, liquidity is also important channel of QE, particularly in the first announcements in QE1 and QE5.

26 See also Bailey et al (2020). The general principle of going big early when QE is used to ease monetary conditions at the ELB has been made by Caldara et al (2020).

27 When the bond market itself is affected by temporary liquidity shortage, as in the “dash for cash” episode in March 2020, it may be optimal to restore market liquidity with central bank tools other than QE (see Hauser (2021)).

All speeches are available online at www.bankofengland.co.uk/news/speeches and @BoE_PressOffice
the extent that it gives an indication of the expected pace of purchases. We could, equivalently, announce a pace directly, an approach taken by, for example, the Federal Reserve and the ECB.

Third, the presence of the ELB puts even greater emphasis on communicating the policy path: when policy rates can move in either direction, we can (even if imperfectly) show what we mean by direct action. If we want to signal that the economy needs stimulus, we cut rates. But at the ELB, expectations can become de-anchored without guidance and QE revealing our reaction function. Clear communications are equally important when considering any unwind of QE. Small, gradual declines in the central bank balance sheet need not have any tightening effect in well-functioning markets, as long as there is clear communication on the desired future policy stance, in order to avoid sending an inadvertent signal that undermines the central bank’s intentions. Without such clear communications, even a small balance sheet reduction can result in a meaningful tightening of the policy stance.

5. Monetary policy space: thinking beyond my MPC remit

Having argued that monetary policy headroom for easing is limited, and likely to remain limited as \( r^* \) is likely to be persistently low, I want to consider how monetary policy space could be increased. As I am coming to the end of my term as an external MPC member, looking ahead means thinking about a period when I am not on the MPC anymore. I will permit myself the luxury of also thinking about policies that are not part of the MPC’s remit, and are decisions for the government to consider, but could nevertheless increase monetary policy space in the future. Let me be very clear here that I am not recommending any particular policy.

Rather, I am providing a list of policies that others (not the MPC) could consider in order to rebuild monetary policy space. These policies are not without risk, but neither is remaining in the status quo with limited monetary policy headroom.

As an organising framework, it is useful to think of monetary policy space for easing as how far we can push real interest rates below neutral, so \( r - r^* \).

In turn, we can decompose the real rate into a nominal and inflation component: \( r = R - \pi^e \).

Total policy space is therefore equal to \( R - \pi^e - r^* \).

The amount of policy space can be increased in three ways: by allowing nominal rates \( (R) \) to go lower, by allowing inflation \( (\pi^e) \) to go higher, and by implementing policies that push \( r^* \) higher.

---

28 The experience of two Fed tightening strategies provides strong evidence for this asymmetry in my view. The 2013 taper tantrum was a decision about the future flows of QE that sent an inadvertent and disruptive signal about the policy rate path. The 2017-2018 QE unwind experience, on the other hand, showed that gradual and predictable balance sheet reduction could take place without any impact on bond risk premia, while rate expectations were managed through both clear communication about the desired pace of tightening, as well as via the actual pace of tightening in policy rates which started to move up slowly from late 2015.
a. Lower nominal rates

The MPC lowered its assessment of the ELB from 0.5% in 2009 and to 0.1% in 2016. In 2021, negative rates were added to the toolbox. I would be comfortable with cutting Bank Rate to -0.5% or even -0.75% the next time monetary stimulus is required. This additional space is helpful, but policy space is still limited relative to pre-crisis years.

And even though I am unambiguously in favour of using negative rates the next time the economic outlook deteriorates and requires more stimulus, I also believe that the economic impact of further rate cuts is likely to decline somewhat at lower levels of interest rates, though the impact would still be positive. This further adds to the argument that, even with negative rates in the toolbox, policy space is limited relative to the pre-GFC years.

The key constraint that simultaneously limits how low the policy rate can go and how effective it is at low levels is that cash is available as an alternative asset, and it pays zero interest.

This zero interest feature of cash is usually described as the cost of holding cash when interest rates are positive. But when interest rates go negative, the zero rate on cash becomes a benefit of holding cash, in effect a subsidy, relative to other assets. The more negative interest rates become, the more attractive cash becomes, potentially leading to a drain on the banking system at some point, of either profits or deposits, or both, which might have a counterproductive effect on the economy at low enough levels of interest rates.

However, neither the cost (when interest rates are positive) nor the subsidy (when interest rates are negative) of cash are intrinsically desirable features. Rather, they represent a technological constraint, namely that it is rather impractical to either pay or charge interest on cash. However, as central banks, including the Bank of England, are considering a move to central bank issued digital currency (CBDC), this constraint can potentially be moved more easily in the future. If digitisation becomes sufficiently widespread so that cash is used much less, this opens up the possibility of having more deeply negative interest rates in the distant future, without causing any negative effect on bank profits since interest rates on all safe assets would become negative, so banks can maintain their net interest margin, as bank deposits are no less attractive than other negative rate assets.

29 Policies to offset this technological constraint do exist, going all the way back at least to Gesell (1906) who recommended taxing money holdings. For an overview of policies to remove or reduce the subsidy on cash, see Rogoff (2017).
30 Cunliffe (2021) documents, among other things, the shrinking importance of cash for everyday transactions. See also Bank of England (2020). There is no plan to abolish cash, and such a plan is not a requirement for lowering the ELB. As Rogoff (2017) points out, it is sufficient to have a “less cash” economy, not a “cash-less” economy.
31 It is important to note that, in the available empirical evidence so far from countries that have implemented moderately negative rates, bank profits have not fallen relative to a counterfactual where interest rates had not been cut into negative territory, as various deposit rates and other funding rates did turn negative, and loan losses were reduced, see Tenreyro (2021) for an extensive review of the evidence.
b. Higher Inflation

The second component of policy space is inflation. Higher inflation would create more monetary policy space. This avenue of creating policy space is restricted by the inflation target, which is set by the government. Low and stable inflation is a very good thing, and the inflation targeting framework has served the UK very well. It is, in any case, not for the MPC to question its own remit. But a number of academics have put forward the argument that the inflation target could be revised. An inflation target will remain crucial as a nominal anchor, but it might be a slightly higher target than today. And the higher target might be permanent, or it might be temporary and conditional on certain circumstances, for example a temporarily higher target when policy rates approach their effective lower bound.

32 For example, Andrade et al (2020) show in a calibrated DSGE model (which accounts for the welfare costs of higher inflation) that in the region where r* has been the optimal inflation target moves almost one for one with r* (i.e. if r* falls by 2 percentage points the inflation target should rise by 2 percentage points).

33 See Bernanke (2017) and Bernanke et al (2019) for discussion.

34 There is a further option here, which is to promote immigration by those who are in a low asset stage of their life-cycle, essentially young workers. This is a potential solution for an individual country, but at the expense of the country from which the immigrants are leaving.

c. Higher r*

The third component of monetary policy space is the level of the neutral rate itself. Like the inflation target, it is taken as given by the MPC. The MPC has no tools or remit to increase it. But thinking beyond the MPC, policies might be available that increase r*. My discussion of the structural drivers of low r* offers a roadmap here.

First, there is demographics. The three key variables here are the birth rate, longevity, and the time spent in retirement.

Pushing up the birth rate simply for the purpose of raising r* seems too radical an option. Changes in the birth rate only have a temporary – though quite persistent – impact on the age profile of the population, but a permanent effect on its growth rate. And higher population growth has other consequences, in particular related to climate change, that are less desirable.

Lowering longevity… well, I am going to rule that out as a policy option for obvious reasons.

That leaves us with reducing the time spent in retirement, which should be on the table. Many countries, including the UK, are already slowly raising their retirement age, though it is by no means keeping up with the increase in longevity. The question is whether it can be increased more, sooner. There are important distributional consequences that need to be taken into account: longevity varies significantly with income, for example. And not all jobs are amenable to be carried out by older workers. Nevertheless, one might at least consider removing any policies that compel workers to retire before they want to, and creating incentives so
that it is financially attractive for those who are able and willing to work, to keep doing so. The increase in the share of part-time work, and the flexibility to work from home, newly boosted by the pandemic, are likely to be helpful to keep older workers in the labour force for longer. The biggest increase in labour force participation over the past 15 years has been in the 50-64 age bracket, and the second biggest has been in the over-65 age bracket. We need more of that.

Next, we should consider income inequality, which can lead to higher debt and higher wealth inequality over time, developments which are, for many, undesirable in their own right, in addition to the fact that they reduce monetary policy space. What causes income inequality, and what policies are available to reduce it?

This is a huge topic, and I am not going to do it justice here. But let me just sketch out two broad policy areas that are relevant: regulatory policy and redistributive taxation.

Over the past several decades, many advanced economies have experienced a rise in firm concentration that followed widespread deregulation. This has led to concerns about reduced competition in the product markets that firms sell into, and reduced competition in the labour markets that firms buy from. Rising profits, weaker investment (but a rise in spending on lobbying and political influence) and a reduction in wages have been the result. Strong regulatory anti-trust legislation and the promotion or facilitation of collective bargaining in labour markets are some of the policies that could help restore the balance of power and therefore the balance of incomes.

A second avenue is taxation, including benefits, which economists often refer to as negative taxes. Falling corporate tax rates and falling marginal income tax rates on the richest, as well as a range of other tax policies, have contributed to rising income inequality since the early 1980s. Flexibility in tax regimes for high income individuals to reclassify income as corporate profits have also played a role. Low inheritance taxes allow income inequality in one generation to become entrenched in future generations.

In both regulatory policy and taxation, there is a pendulum that has swung from a high tax, high regulation environment to a low tax, low regulation environment since the early 1980s. To some extent that reflected genuine concerns about a stifling environment that impeded growth for all. But one might reasonably argue

25 The UK has already moved significantly in this direction, for example scrapping the Default Retirement Age in 2011. The question is whether more can be done to raise the average effective retirement age.
27 See Farber et al (2021) for the effect if unions on inequality and Philippon and Reshef (2012) for the effect of deregulation on income earned in the finance sector.
28 A general conclusion from the literature on inequality is that capital income is key in explaining inequality at the top of the income and wealth distribution (including ways richer individuals are able to have their income and capital gains at lower rates), and that income and wealth of the richest are likely to be substantially understated due to tax evasion and wealth held offshore (the latter being significant for the UK). See Dabla-Norris et al (2015) and OECD (2014) for trends in global wealth inequality and marginal tax rates; Terslav et al (2020) for the effect of corporate profit shifting; Aistadsater et al (2018, 2019) and Zucman (2019) for the effect of offshore wealth and evasion on wealth inequality. Clark and Leicester (2005) discuss specific UK tax and benefit changes and their impact on income inequality. Corlett et al (2020) discuss the importance of UK capital gains in accurately measuring the incomes of the highest earning UK households.
that the pendulum has swung too far, and that the low tax, low regulation regime combined with globalisation ended up widening income disparities in a way that not only hurt those at the bottom of the income distribution, but ended up having adverse macroeconomic effects: an undesirable high debt and low productivity growth environment that is simultaneously more fragile and reduces the policy space to fix it when it threatens to break.

It is not for central bankers to decide any of these measures. But it is for central bankers to point out that this is the fragile macroeconomic environment we have ended up in, and that (non-central bank) policies exist to improve the situation. None of the options for change are easy or free of risk, but they need to be judged against an increasingly risky and untenable status quo.

The time to have the debate on how to restore policy room is now, in a recovering economy. It would be a mistake to wait until the next downturn.

6. Conclusion

I have presented new theoretical and empirical insights from the past few years to support my argument that demographics, debt and income inequality are important factors in lowering the neutral rate of interest.

I argue that we are only about two thirds of the way through a multi-decade demographic transition that is affecting interest rates. Absent policy changes, there is no prospective reversal in this particular driver of interest rates: downward pressure from demographics either continues further or remains where it is. The key mechanism is not that older people have lower savings rates, but rather that, as people age, they hold higher levels of assets (the accumulated stock savings over their lifetime), in particular safe assets, and those assets are only run down slowly and partially late in life. The higher saving of the middle-aged outweighs the modest dissaving of the retirees.

I also summarise some new research that links debt, income inequality and wealth inequality. Recent research argues that the fundamental driver is income inequality, and higher debt and wealth inequality follow from it. Higher income households want to accumulate assets, including lending to lower income households. This increases private debt and lowers interest rates by acting as a credit supply shock, by making economic outturns riskier due to a higher probability and severity of financial crises, and possibly by weighing on future consumption demand of indebted households.

See Philippon (2019) and Saez and Zucman (2019).

A notably positive development to counter this fragility has been the significant regulatory reforms in the financial sector in the aftermath of the financial crisis. The UK and global banking system has far more capital and liquidity now. High debt will still weigh on demand growth, especially in a recession, but it is less likely to be amplified via a reduced ability of the banking system to continue supplying finance.

Another positive development is the recent global policy initiative to raise the minimum corporate tax rate and reduce profit shifting to tax havens.
The policy implications are even more stark than when I first discussed them nearly six years ago. We have limited headroom for easing monetary policy, so we will not be able to provide monetary stimulus on the same scale as in previous recessions. QE headroom, which should be measured in basis points of yields rather than in billions of bonds available for purchase, is limited as well.

To address limited monetary easing space, there are three types of policy available. Changes that enable policy rates to be cut into deeply negative territory; temporarily or permanently higher inflation rates; policies that raise the neutral rate by lowering time spent in retirement or lowering inequality. None of these policies are within the MPC’s remit. These are policies for government to decide. I am merely pointing out that, without further action, we will remain stuck with limited headroom for monetary easing.

This speech has not been about the near-term outlook for the economy. On this, I will be brief. First, even though the expected peak in inflation now looks to be higher than previously expected, I have not changed my view that this inflation peak is likely to be temporary. It is driven by supply bottlenecks and base effects, both of which are set to wane next year. Second, we are not out of the woods yet in terms of the virus and the impact on the economy. Yes, the economy has been growing rapidly, but on the most recent data it remains an average recession away from full employment. Monthly GDP in May was 4 ½ % below its December 2019 pre-Covid level, the unemployment rate in May was 1pp (about 300,000 people) above its pre-Covid level, and 1.3 million jobs remained fully or partially on furlough at the end of June. The delta variant is still causing health and economic damage, both in the UK and in the rest of the world, in a way that risks feeding back to the UK, economically. Third, various government support schemes are coming to an end, including the all-important furlough scheme. I would want to see how the economy copes with that, before adding monetary tightening on top of fiscal tightening. For all these reasons, I think it will remain appropriate to keep the current monetary stimulus in place for several quarters at least, and probably longer. And when tightening does become appropriate, I suspect not much of it will be needed, given the low level of the neutral rate.
References


All speeches are available online at www.bankofengland.co.uk/news/speeches and @BoE_PressOffice


Froemel, M., M. Joyce, and I. Kaminska (2021), ‘What to expect when they’re expecting’, Bank Underground article available at [https://bankunderground.co.uk/2021/03/26/what-to-expect-when-theyre-expecting/](https://bankunderground.co.uk/2021/03/26/what-to-expect-when-theyre-expecting/)


Rannenberg, A. (2021), ‘Inequality, the risk of secular stagnation and the increase in household debt’, NBB Working Papers, No. 375.


All speeches are available online at www.bankofengland.co.uk/news/speeches and @BoE_PressOffice

21