The State of New Keynesian Economics: A Partial Assessment

Jordi Galí*

CREI, Universitat Pompeu Fabra and Barcelona GSE

October 2017

Preliminary. Do not circulate.

* Correspondence: Ramon Trias Fargas 25; 08005 Barcelona (Spain). Email: jgali@crei.cat. I thank Christian Höynck and Cristina Manea for excellent research assistance.
In August 2007, when the first signs emerged of what would come to be the most damaging global financial crisis since the Great Depression, the New Keynesian paradigm was dominant in macroeconomics. The New Keynesian model (henceforth, the NK model), was taught in economics programs all over the world as the framework of reference for understanding fluctuations in economic activity and inflation, and their relation to monetary and fiscal policies. It was widely adopted by researchers as a baseline model that could be used flexibly to analyze a variety of macroeconomic phenomena. The NK model was also at the core of the estimated medium scale dynamic, stochastic, general equilibrium (DSGE) models developed by central banks and policy institutions throughout the world.

Ten years later little has changed: the NK model arguably remains the dominant framework in the classroom, in academic research, and in policy modeling. No alternative paradigm has emerged that can claim success, even partial, at replacing the NK model. This is true despite the tons of ammunition fired against modern macroeconomics in general, and against DSGE incarnations of NK models in particular, with criticisms focusing on their failure to predict the crisis and to provide useful guidance to policymakers in their attempts to stimulate the economy.

In fact, one may argue that over the past ten years the scope of New Keynesian economics has kept widening, by encompassing a growing number of phenomena that are analyzed using its basic framework. The main objective of the present paper is to take stock of the state of New Keynesian economics, by reviewing some of its main insights and by providing an overview of some of recent developments. In particular, I discuss some recent work on two very active research programs: the implications of the zero lower bound (ZLB, henceforth) on nominal interest rates and the interaction of monetary policy and household heterogeneity. Finally, I discuss what I view as some of the main shortcomings of the NK model and possible areas for future research.

The New Keynesian Model: A Refresher

Modern New Keynesian economics can be interpreted as an effort to combine the methodological tools developed by Real Business Cycle (RBC) theory with some of the central tenets of Keynesian economics tracing back to the General Theory.

A hallmark of the approach to modeling economic fluctuations pioneered by RBC theorists is the reliance on dynamic, stochastic, general equilibrium frameworks. In practice, the latter take the form of a set of stochastic difference equations that describe, in a highly aggregative manner, (i) the behavior of households, firms and policymakers, (ii) some market clearing and/or resource constraints, and (iii) the evolution of one or more exogenous variables that are the ultimate source of fluctuations in the economy. All the previous features seem natural ingredients of any model that seeks to explain
economic fluctuations, and as such they have been fully adopted by New Keynesian economics.¹ Less uncontroversial may be the assumption, widely found in both RBC and NK models, that the behavior of households and firms (and, in some instances, of policy makers as well) is the outcome of an optimization problem, solved under the assumption of rational expectations, though a strand of the recent literature (not reviewed here) has examined the consequences of relaxing the latter assumption.

What does New Keynesian economics add to the standard RBC apparatus? One can pinpoint three modifications which, combined, generate a substantially different framework. Firstly, it introduces nominal variables (i.e. prices, wages, the nominal interest rate, etc.) explicitly. Secondly, it departs from the assumption of perfect competition in the goods market, leading to positive price markups. Thirdly, it introduces nominal rigidities, typically in the form of constraints on the frequency with which prices can be adjusted.² The assumption of imperfect competition is often extended to the labor market as well, with the introduction of wage rigidities (nominal or real).

The resulting framework implies a breakdown of monetary policy neutrality. This has several consequences. Firstly, exogenous changes in monetary policy have non-trivial effects on real variables, not only on nominal ones. In addition, and more importantly, the economy’s equilibrium response to any shock is not independent of the monetary policy rule in place, thus opening the door to a meaningful normative analysis of alternative policy rules.

It is convenient, for future reference, to lay out the two equations that describe the non-policy block of the NK model, in its simplest version (with sticky prices but flexible wages):

\[ \tilde{y}_t = E_t \{ \tilde{y}^{s+1}_t \} - \sigma^{-1} (i_t - E_t \{ \pi_t \} - r^n_t) \]  

(1)

\[ \pi_t = \beta E_t \{ \pi_t^{s+1} \} + \kappa \tilde{y}_t \]  

(2)

where \( \tilde{y}_t \) is the output gap (given by the difference between log output \( y_t \) and log natural output \( y^n_t \)), \( i_t \) is the nominal rate, \( \pi_t \) denotes inflation, and \( r^n_t \) is the natural rate of interest. Natural output and the natural rate of interest are the values that those variables would take in equilibrium if prices were fully flexible. Equation (1), often referred to as the dynamic IS equation, implies that the output gap \( \tilde{y}_t \) differs from its expected value one period ahead by an amount that is proportional to the gap between the real interest rate and its natural counterpart. It can be derived by combining the Euler equation describing the optimal consumption behavior of the representative

---

¹ To put it differently: It is only too easy to imagine the criticisms that would be aimed at modern macro if the latter relied on static, deterministic or partial equilibrium models.

² Nominal rigidities are generally introduced using the formalism proposed by Calvo (1983), whereby a constant fraction of firms, drawn randomly from the population, are allowed to adjust the price of their good.
household with a goods market clearing condition requiring that output must equal consumption. Equation (2), the so-called *New Keynesian Phillips curve*, states that inflation depends on expected inflation one period ahead and the output gap. It can be derived in two stages. In a first stage, a relation between inflation, expected inflation and the markup gap (i.e. the log deviation of average markup from the desired markup) can be derived by aggregating the optimal price setting decisions of firms subject to constraints on the frequency with which they can adjust prices. That relation is combined with labor supply equation, a goods market clearing condition and an aggregate production function to obtain a simple relation linking the markup gap to the output gap, thus giving rise to (2).³

In order to close the model, a description of monetary policy (i.e. of how the nominal rate \( i_t \) is determined) is needed. A Taylor-type rule of the form

\[
i_t = \phi_i \pi_t + \phi_y \hat{y}_t + \nu_t
\]

where \( \hat{y}_t \) denotes the log deviation of output from steady state, and \( \nu_t \) is an exogenous monetary policy shifter following some stochastic process, is an example of such a description, and one frequently used in the literature as an approximation to the conduct of monetary policy in advanced economies (at least in normal times).⁴

Despite its simplicity, the basic NK model described above yields several interesting insights, some of which I briefly discuss next.

First, the model implies that monetary policy is not neutral. Non-neutrality has (at least) two dimensions. A first dimension has to do with the fact that an exogenous monetary policy shock affects not only nominal variables, but also real ones (like output). In particular, a persistent increase in \( \nu_t \) (i.e. an exogenous tightening of monetary policy) raises both nominal and real rates, leading to a fall in output and inflation, in a way consistent with the evidence.⁵ The second dimension of non neutrality has to do with the fact that the response of output (and other real variables) to a non monetary shock (e.g. a shock that changes the natural levels of output and/or of the interest rate, \( y^*_t \) and \( r^*_t \)) is not invariant to the monetary policy rule adopted by the central bank.

Interestingly, when interest rate rule (3) is calibrated in a way consistent with U.S. evidence for the post-1982 period, the model implies responses to technology shocks

---

³ See Woodford (2003) or Gali (2015) for a detailed derivation of (1) and (2) and a discussion of the underlying assumptions.

⁴ See, e.g. Taylor (1993).

⁵ See, e.g., Gali (2015) for the implied response of a calibrated NK model to both monetary and technology shocks. See, e.g., Christiano et al. (1999) for empirical evidence on the effects of monetary policy shocks.
consistent with the empirical evidence, including a countercyclical response of employment.\textsuperscript{6}

Other insights generated by the analysis of the NK model pertain to its normative implications for the conduct of monetary policy. A key finding in that regard has to do with the conditions on interest rate rule (3) which guarantee a unique equilibrium (locally, i.e. in a neighborhood of the steady state). As shown by Bullard and Mitra (2002) that condition takes the form

\[ \kappa(\phi_x - 1) + (1 - \beta)\phi_y > 0 \]

which requires that the central bank adjusts the policy rate sufficiently strongly in response to variations in inflation and output, a condition known as the Taylor principle. If that condition is not satisfied the equilibrium is locally indeterminate, opening the door to fluctuations driven by self-fulfilling revisions in expectations (a.k.a. sunspot fluctuations).\textsuperscript{7}

Beyond simple rules like (3), the literature has sought to characterize the optimal monetary policy, i.e. a policy that minimizes the representative household’s welfare losses. The latter can be approximated by the function:

\[
E_0 \sum_{t=0}^{\infty} \beta^t \left( \pi_t^2 + \Theta(y_t - u_t)^2 - \Phi \hat{y}_t \right)
\]

where \( \hat{y}_t \equiv y_t - y \) denotes the (log) deviation of output from its zero inflation steady state, and \( u_t \equiv \hat{y}_t^e - \hat{y}_t^n \) represents the (possibly time-varying) gap between the efficient and the natural levels of output, \( \hat{y}_t^e \) and \( \hat{y}_t^n \), both expressed in logs and deviations from their respective steady state values. Thus, the term \( \hat{y}_t - u_t \) has the interpretation of fluctuations in the gap between output and its efficient level. Such deviations are a source of welfare losses. In addition, deviations from zero inflation generate losses due to the misallocation of resources caused by the associated price dispersion. Finally, the term \( \Phi \hat{y}_t \) captures the welfare gains from positive deviations of output from steady state when the latter is associated with an inefficiently low level of activity, with coefficient \( \Phi \) being proportional to the extent of that inefficiency.

In the special case of an efficient natural level of output (implying \( \Phi=0 \), and \( u_t = 0 \) for all \( t \)), welfare losses can be minimized by fully stabilizing inflation at zero (i.e. \( \pi_t = 0 \) for all \( t \)). Note that the NK Phillips curve (2) implies that such strict inflation targeting policy has an important byproduct: it stabilizes the output gap at zero, thus making output equal to its natural (and, by assumption, efficient) level, a property is sometimes

\textsuperscript{6} See, e.g. Galí (1999) and Basu et al. (2006) for evidence on the U.S. economy’s response to identified technology shocks.

\textsuperscript{7} See Clarida et al. (2000) for empirical evidence suggesting that the local uniqueness condition may not have been satisfied during the pre-Volcker era, potentially giving rise to unnecessary instability.
referred to as the Divine Coincidence.\textsuperscript{8} As a result welfare losses are zero and the economy attains its first-best allocation.

The previous result holds only when the flexible price equilibrium allocation is optimal, i.e. when nominal rigidities are the only distortion in the economy. More generally, the presence of real frictions is likely to drive a wedge between the natural and efficient levels of output. This is reflected in a positive $\Phi$ (in the case of an inefficient steady state) and/or in non-trivial fluctuations in $u_t$ (when the presence of real frictions implies an inefficient response of natural output to some shocks). As a result, a trade-off emerges between price stability and the attainment of an efficient level of economic activity, thus giving rise to a nontrivial optimal monetary policy problem.\textsuperscript{9} The optimal policy along with its associated output gap and inflation outcomes depend on one’s assumption regarding the extent to which the central bank can credibly commit to a state-contingent plan. Standard treatments of the optimal monetary policy problem and its consequences have focused on the extreme cases of full discretion (i.e. period-by-period re-optimization) and full commitment (a once-and-for-all choice of an optimal plan, which is subsequently followed through, even though it may be suboptimal ex-post to do so). An analysis of some intermediate cases can also be found in the literature.\textsuperscript{10}

The study of the optimal policy in the context of the NK model has yielded several interesting insights. Among the latter is the nature of the gains from commitment and the kind of inefficient outcomes or biases implied by discretionary policies. Thus, the presence of an inefficiently low steady state output (reflected in $\Phi>0$), combined with the lack of commitment, generates a (suboptimal) positive inflation bias, similar to that uncovered by Kydland and Prescott (1980) and Barro and Gordon (1983) in the context of an earlier generation of monetary models with non-neutralities. Most interestingly, even when the steady state is efficient ($\Phi=0$), gains from commitment arise in the presence of shocks that imply an inefficient response of natural output (i.e. fluctuations in $u_t$, even if the latter has a zero mean), as it would be the case in the presence of certain real imperfections. Those gains result from the forward-looking nature of inflation in the NK model, combined with the central bank’s ability to influence expectations of future inflation and output gaps, which makes it possible to smooth over time the deviations from the first best allocation, thus reducing the implied losses. By contrast, in the absence of commitment, the central bank has to rely exclusively on its ability to affect the current output gap, which leads to excessive fluctuations in both inflation and the gap between output and its efficient level (i.e. the gap measured by $\tilde{y}_t - y_t$) and, hence, to larger welfare losses. The resulting excess volatility associated with the discretionary policy is sometimes referred to as stabilization bias, and may

\textsuperscript{8} See Blanchard and Gali (2005) for a discussion of the Divine Coincidence
\textsuperscript{9} See Clarida et al. (1999) for an analysis and discussion of the optimal monetary policy problem under discretion and under commitment.
\textsuperscript{10} See Schamburg and Tambalotti (2007) and Debortoli and Lakdawala (2016)
coexist with an optimal \textit{average} level of inflation (in contrast with the case of an inflation bias).

While the notion of a once-and-for-all choice of an optimal state contingent plan is utterly unrealistically, the analysis of the optimal policy under commitment establishes a useful benchmark that can be used to inform the search for simple rules that can approximate that policy. Thus, the analysis of the properties of the equilibrium under the optimal policy with commitment, which often imply a stationary price level, provides a possible rationale for the adoption of a price level targeting interest rate rule, of the form \( i_t = \phi_p (p_t - p^*) \). As shown by Vestin (2006), the latter rule can approximate well the optimal policy with commitment, given a suitable choice of coefficient \( \phi_p \).

Many other interesting insights regarding the optimal design of monetary policy have emerged from the analyses of relatively straightforward extensions of the basic NK model described above, allowing for staggered wage setting (e.g. Erceg et al. (2000)), some backward-looking price setting (Steinsson (2003)), open economy considerations (e.g. Clarida et al. (2002), Galí and Monacelli (2005)), deviations from rational expectations (e.g. Evans and Honkapohja (2003), Woodford (2010)), labor market frictions (Trigari (2009), Blanchard and Galí (2010)), uncertainty shocks (Basu and Bundick (2017)), among many other extensions.

The next two sections focus on two specific extensions of the basic NK model that have drawn considerable attention in recent years and triggered a good amount of research: the zero lower bound on the nominal interest rate and households’ heterogeneity.

\section*{The Zero Lower Bound}

The nimble response of central banks to the recessionary and deflationary forces triggered by the financial crisis was seemingly jeopardized when, after being successively reduces, policy rates attained the lower bound of zero (or close to zero).\footnote{The zero lower bound on nominal interest rates is a consequence of the existence of cash as an alternative investment with a zero nominal return.} The basic NK model, described in the previous section, ignores the existence of the zero lower bound (ZLB). A number of papers, originally motivated by the Japanese experience with a liquidity trap, adopted the NK framework to analyze the implications of a binding ZLB.

To illustrate some of the insights of that analysis let us consider the case of an efficient natural equilibrium (i.e. \( u_t = 0 \), for all \( t \)). In the absence of the ZLB constraint, the optimal policy implies full stabilization of the output gap and inflation, as discussed above. Let us augment equations (1) and (2) with the ZLB constraint...
for all $t$. Consider next a one-off episode with a temporary, but persistent adverse shock to the natural rate, $r''$, that brings the latter into negative territory. It should be clear just by looking at equations (1) and (2) that attaining the first best outcome is no longer feasible in the presence of the ZLB constraint. As implied by (1), the inability to match the drop in the natural rate of interest with a commensurate reduction in the policy rate generates, for any given path of inflation, a persistent negative output gap. The fact that the latter is itself a source of persistent deflation, as implied by (2), leads to a higher real interest rate and, thus, to an even larger gap between that variable and its natural counterpart, deepening further the initial recession. An analysis of the optimal design of monetary policy in the presence of a ZLB on the nominal rate closely related to the exercise above can be found in Jung, Teranishi and Watanabe (2003) and Eggertsson and Woodford (2003). Both papers study the case of a fully unanticipated, once-and-for-all adverse shock to the natural rates, which pushes the optimizing central bank against the ZLB.

Figure 1, based on the analysis in Galí (2015), illustrates some of the implications of the ZLB for the conduct of monetary policy. It simulates the response to an unanticipated demand shock that lowers the natural rate of interest from its normal steady state level of 4% to -4% (both in annual terms) between period 0 through 5. In period 6 the natural rate goes back to its initial value, something which is assumed to be (correctly) anticipated as of period 0, when the shock hits. In the absence of a ZLB, price stability and a zero output gap could be maintained in the face of the adverse disturbance, if only the central bank were to lower the interest rate to -4% during the duration of the shock, thus tracking the path of the natural rate. The existence of the ZLB, however, makes that option unfeasible. Once again, the nature of the optimal policy depends on the extent to which the central bank can commit to future actions. The line with circles plots the response of the output gap, inflation and the nominal interest rate under the optimal discretionary policy (i.e. without commitment). Note that, in response to the adverse shock, the central bank lowers the nominal rate to zero and keeps it there until the shock goes away, in which case it brings it back to its initial level (4%) consistent with price stability. Both output and inflation experience large declines in response to the shock and take persistent negative values until the adverse disturbance vanishes, when the central bank can fully restore price stability and close the output gap. The line with crosses displays the equilibrium responses under the optimal policy with commitment. Now the central bank credibly promises that it will keep the nominal rate at zero even after the shock is no longer effective (for two periods longer in the simulated example). That policy leads to a small, but non-zero, output gap and inflation in subsequent periods, implying a welfare loss relative to the first best. But that loss is more than offset by the gains resulting from the much greater stability in earlier periods, when the disturbance is active. That optimal policy with commitment can be interpreted as an illustration of the power of forward guidance policies, i.e. policies that aim at influencing current macro outcomes through the management of expectations about...
future policy settings. Such policies have been openly adopted by central banks like the ECB and the Federal Reserve in the aftermath of the Great Recession, and in the face of a very slow recovery.

The previous example illustrates the monetary policy implications of a fully unanticipated, one-off, temporary drop in the natural rate of interest to a negative level. A number of authors have instead analyzed an economy where the natural rate of interest is subject to recurrent shocks. In those economies, the possibility of hitting the ZLB constraint in the future affects how the economy responds to shocks (and to policy) even when the ZLB is not binding. Adam and Billi (2006, 2007) and Nakov (2008) study the implications of the ZLB for the optimal design of monetary policy, in a stochastic setting, when that constraint is occasionally (but recurrently) binding. Several insights emerge from that analysis. First, the optimal policy implies a nonlinear response to shocks, with the central bank reducing nominal rates more aggressively in response to adverse shocks, in order to reduce the probability of a binding ZLB down the road. Furthermore, under commitment, the optimal policy calls for sustained monetary easing even when the natural rate is no longer negative. Secondly, the gains from commitment (relative to discretion) are much larger than in the absence of a ZLB. Finally, as stressed by Nakov (2008) a large fraction of the gains from commitment can be reaped by adopting a simple price level targeting rule, which also reduces the incidence of a binding ZLB considerably.

A recent paper by Nakata and Schmidt (2016) provides a new rationale, connected to the ZLB, for Rogoff’s (1985) case for appointing a "conservative" central banker (i.e. one that puts more weight than society on inflation stabilization), even in the absence of the conventional inflation bias. They show that, under an optimal discretionary policy, the anticipation of an occasionally binding ZLB implies that on average inflation falls below target and the output gap is positive, even when the ZLB is not binding. Delegating monetary policy to a “conservative” central banker is generally desirable since the latter will keep inflation closer to target (at the cost of an even larger output gap) when the ZLB is not binding, with the anticipation of that policy providing a highly welcome additional stimulus when the ZLB is binding, and improving social welfare.

*The Forward Guidance Puzzle*

The high effectiveness of forward guidance policies illustrated by the example above is closely linked to a property of the NK model whose empirical relevance has been called into question by several authors, giving rise to the so called *forward guidance puzzle*.\(^{12}\) It is useful, in order to describe that puzzle, to iterate forward the dynamic IS equation

---

\(^{12}\) Carlstrom et al. (2015) and Del Negro et al. (2015) were the first papers to (independently) point out the forward guidance puzzle (the label was given by the latter).
and, under the assumption that the output gap is expected to converge to zero asymptotically, to rewrite it as:

$$\tilde{y}_t = -\sigma^{-1} \sum_{t=0}^{\infty} E_t \{ i_{t+k} - \pi_{t+k+1} - r^u_t \}$$

(5)

To understand the implications of the previous relation, let us assume first that prices are fully rigid so that $\pi_t = 0$ for all $t$. The forward guidance puzzle can be stated as follows: the effect on output of an anticipated change in the policy rate of a given size and duration is independent of the timing of its implementation. In other words, the effects of a temporary 1% increase in the policy rate 100 years from now are predicted to be the same as if the increase were to take place immediately or in the near future. The reason behind that prediction is that the dynamic IS equation (1) implies no discounting of the expected output gap and, as a result, (5) involves no discounting of future interest rates. The previous puzzle is amplified once we relax the assumption of fully rigid prices and let inflation be determined by (2) instead. In that case, the longer is the horizon of implementation of a given change in the policy rate, the stronger are its effects on output, and hence the larger and more persistent the response of inflation. For any given path of the nominal rate, the latter effect works in the direction of changing the real interest rate in a way that further amplifies the effects on output and inflation, leading to a strong nonlinear effect due to the accumulation of feedback effects.

Several authors have proposed “solutions” to the forward guidance puzzle. They generally involve modifications of the benchmark NK model that generate some kind of discounting in the consumer’s Euler equation. Examples of such modifications include the introduction of finite lives (Del Negro et al. (2015)), incomplete markets with (Farhi and Werning (2016)) or without (McKay et al. (2016, 2017)) bounded rationality, lack of common knowledge (Angeletos and Lian (2017)), and behavioral discounting (Gabaix (2017)).

The resulting dynamic IS equation typically takes the form

$$\tilde{y}_t = \alpha E_t \{ \tilde{y}_{t+1} \} - \sigma^{-1} (i_t - E_t \{ \pi_{t+1} \} - r^u_t)$$

(6)

where $0 < \alpha < 1$. Iterating forward, one can derive the following expression for the current output gap:

$$\tilde{y}_t = -\sigma^{-1} \sum_{t=0}^{\infty} \alpha^t E_t \{ i_{t+k} - \pi_{t+k+1} - r^u_{t+k} \}$$

(7)

which implies that the effects on current output of anticipated changes in the real interest rate decline with the horizon of implementation. As long as prices are not fully rigid, however, the previous feature reduces the effect on inflation of any anticipated change in the real rate, but it does not overturn the predictions that the size of such an effect increases with the horizon of implementation.
Global Multiplicity and the ZLB

Much of the analysis based on the NK model found in the literature has a local nature and is carried out using a linear approximation to the equilibrium conditions around a steady state consistent with the inflation target (typically zero). Several papers, however, have explored the properties of equilibria of the NK model from a global perspective. The work of Benhabib et al. (2001) triggered much of the research on this front by showing that a Taylor-type interest rate rule satisfying the ZLB constraint and consistent with a locally unique equilibrium around the steady state associated with the targeted inflation rate necessarily implies the existence of another steady state, labeled as a liquidity trap, in which the interest rate is zero or near zero and inflation is below the targeted level and possibly negative. Furthermore, and more worrisomely, there exist an infinite number of equilibrium trajectories that converge to the liquidity trap steady state. In a companion paper, Benhabib et al. (2002) point to possible solutions to the global multiplicity problem, while preserving the desirable local properties of an active Taylor rule in a neighborhood of the intended steady state. The solutions proposed include the activation of alternative monetary and fiscal rules in case the economy enters a path leading to the liquidity trap steady state. The proposed rules feature in one case a strong fiscal stimulus in the form of lower taxes, and a switch to a money growth rate peg in the second case. Under those rules, any path converging to the liquidity trap would violate the intertemporal budget constraints of the government and households, and can thus be ruled out as an equilibrium path.13

Several papers have provided “quantitative” applications of the global equilibrium multiplicity implied by the ZLB in the NK model. Schmitt-Grohé and Uribe (2017), Aruoba et al. (2017) and Jarocinski and Mackowiak (2017) use a quantitative NK model with global multiplicity to interpret the prolonged recession and persistently low inflation in many advanced economies in the wake of their financial crises, and despite a highly expansionary monetary policy, with the interest rate at zero. Those papers interpret the crises and subsequent persistent slump with an equilibrium path converging to a liquidity trap steady state. Schmitt-Grohé and Uribe (2017) and Jarocinski and Mackowiak (2017) propose a policy to exit the liquidity trap, consisting of an exogenous path for the policy rate converging to its value in the intended steady state. Under the equilibrium dynamics implied by the liquidity trap, that policy is shown to raise inflation expectations and to stimulate aggregate demand and output. The price level indeterminacy implied by the exogenous path for the nominal rate can be eliminated by a switch to an active fiscal policy. Benigno and Fornaro (2017) develop a model downward nominal wage rigidities and a ZLB constraint similar to that of

13 Cochrane (2011) criticizes the previous approach to ruling out equilibria that deviate from the intended steady state (including equilibria involving hyperinflations). Instead he proposes the specification of policies that are consistent with a unique equilibrium that remains well defined, near or farther away from the intended steady state. A non-Ricardian fiscal policy, combined with a passive monetary policy is an example of an alternative fiscal-monetary regime that avoids the problems of global multiplicity of the NK model with an active Taylor rule.
Schmitt-Grohé and Uribe (2017), but in which they embed an endogenous growth mechanism. Under some conditions, two different balanced growth paths may be consistent with equilibrium. One of those paths, which they refer to as a stagnation trap, is characterized by involuntary unemployment and low growth, while the other features high growth and full employment. Expectations about future growth prospects determine which equilibrium obtains.

In the references above, monetary policy is described by some (generally sub-optimal) Taylor-type rule. But the multiplicity of equilibria generated by the ZLB is not restricted to that case: as shown in Armenter (2016) and Nakata and Schmidt (2016) it also emerges under the assumption of a central bank optimizing under discretion. Even if such multiplicity is clearly suboptimal (since it is associated with inefficient fluctuations) there is little that a central bank can do about it, no matter how good its intentions, since the ZLB limits its ability to stabilize inflation and, as result, may find it optimal in some circumstances to accommodate revisions in private sector’s expectations.

**Fiscal Policy and the ZLB**

In addition to its impact on the design of monetary policy, the ZLB constraint also has implications on the effects of other shocks, including fiscal policy shocks. This is a consequence of a fairly general principle: in the presence of nominal rigidities the effects of any fiscal policy intervention are not invariant to the monetary policy rule in place and, more precisely, to the (endogenous) nominal and real interest rate response to those interventions. By shaping that response, the presence of a ZLB constraint has an impact on the effects of fiscal policy shocks.

The work of Eggertsson (2011) and Christiano et al. (2011) provides an analysis of the interaction of fiscal policy and the ZLB using an NK model as a reference framework. In particular, Eggertsson (2011) considers an environment in which an adverse demand shock pushes the natural rate into negative territory and makes the ZLB binding. In that context, he shows that a reduction in taxes on labor or capital income is contractionary, whereas an increase in government purchases has a strong expansionary effect on output. The reason is that tax cuts (as well as other supply side policies) generate disinflationary pressures that are not matched by a policy rate cut, leading to an increase in the real rate and a drop in aggregate demand. On the other hand, an increase in government purchases has a stronger expansionary effect under a binding ZLB than in “normal” times, since the inflationary pressures generated by the fiscal expansion, combined with the absence of a nominal rate adjustment, lead to a drop in the real rate, thus amplifying the effect of the fiscal stimulus.

Christiano et al. (2011) analyze the determinants of the size of the government spending multiplier in connection with a binding ZLB. In particular, they show that the multiplier is very sensitive to how long the ZLB is expected to bind. When they extend the basic
NK model to allow for endogenous capital accumulation the size of the spending multiplier becomes even larger, since investment—which is inversely related to the real interest rate—responds procyclically to the fiscal shock thus amplifying the effect of the latter. Using an estimated DSGE model they quantify the value of the fiscal multiplier under a binding ZLB to be in a neighborhood of 2 for an increase in spending lasting 12 quarters. This contrasts with a multiplier smaller than one when the model is simulated under “normal” times, with the central bank responding to a fiscal expansion according to a conventional Taylor rule.

The effectiveness of different fiscal policies at stimulating the economy under a binding ZLB is not invariant to the reason why the ZLB has become binding. In the papers discussed above, the ZLB becomes binding as a result of an adverse fundamental shock that is sufficiently large to push the policy rate against the ZLB constraint, making it impossible for the central bank to stabilize inflation and the output gap. Mertens and Ravn (2014) focus instead on expectational or non-fundamental liquidity traps, which may emerge as a result of self-fulfilling expectations, due to the global indeterminacy discussed in the previous subsection. In the case of an expectational liquidity trap, they show that an increase in government purchases has a small effect on output (smaller than in normal times), whereas a tax cut is contractionary. The main factor behind those predictions is the differential effect on inflation: positive in the case of a tax cut, negative for a spending increase. Since the previous predictions are exactly the opposite to those arising in the case of a fundamental liquidity trap, it follows that a good diagnosis of the nature of a liquidity trap is essential in order to evaluate the effects of a given fiscal policy response.

**Heterogeneity**

The standard NK model, like most of its predecessors in the RBC literature, represents an economy inhabited by an infinitely-lived representative household. That assumption, while obviously unrealistic, may be justified by the belief that, like so many other aspects of reality, the finiteness of life and the observed heterogeneity of individuals along many dimensions (education, wealth, income, preference for leisure,...) can be safely ignored for the purposes of explaining aggregate fluctuations and their interaction with monetary policy, with the consequent advantages in terms of tractability.

Yet, as a large literature has made clear by now, the representative household assumption is less innocuous than may appear at first thought, even for the issues that have been the focus of analyses using the NK model. An important reason for this (though not the only one) is the impossibility of studying the consequences of financial frictions of any sort in that framework, since the assumption of a representative household implies that no financial transactions take place in equilibrium and, as a result, any related frictions or imperfections become nonbinding. Thus, in order to
understand whether the presence and nature of financial frictions have non-trivial implications for economic fluctuations and monetary policy it is necessary to relax the representative household assumption. This is precisely what a large (and growing) number of papers have undertaken in recent years, using a suitably modified NK model as a reference framework.

Given space restrictions, I will focus my discussion on the latest generation of models with heterogeneous agents and financial frictions, generally referred to as HANK (short for Heterogeneous Agent New Keynesian models). A key feature shared by the recent wave of HANK models, and which differentiates them from the baseline New Keynesian model is the assumption of idiosyncratic shocks to households’ labor productivity and, hence, to their wage. Those shocks are often assumed to follow a stochastic process that is consistent with some features of the micro data. Secondly, it is generally assumed that only a small number of assets can be traded, as well as the existence of an exogenous borrowing limit. As a result, households cannot perfectly insure themselves against idiosyncratic risk. Furthermore, a (time-varying) fraction of households are against a binding borrowing constraint, which makes their consumption respond strongly to fluctuations in current income. The previous features imply that there is no simple Euler-type equation describing the dynamics of aggregate consumption and, hence, no simple dynamic IS equation like (1) can be derived. On the other hand, the supply side of the model and the description of monetary policy are similar to those in the basic NK framework (as in (2) and (3)).

One of the main lessons emerging from the analysis of HANK models can be summarized as follows: The presence of uninsurable idiosyncratic shocks, combined with the existence of borrowing limits, implies that different households, even if ex-ante identical, may have at any point in time very different marginal propensities to consume (MPCs). As a result, the macroeconomic effects of any aggregate shock will be amplified or dampened depending on the way the shock (and the changes that it triggers) affects the distribution of income and wealth across households.

Several recent papers provide an insightful analysis of that mechanism. Auclert (2017) provides a careful study of the different channels through which heterogeneity shapes the impact of an exogenous monetary policy shock on individual and aggregate consumption. Two of those channels are already present in the representative agent NK model (henceforth, RANK, for short): (i) intertemporal substitution, in response to changes in real interest rates, and (ii) the change in consumption induced by the resulting changes in aggregate income, which is a source of a multiplier effect. A HANK economy, on the other hand, provides additional channels, that are a consequence of the redistribution that takes place in response to a monetary policy

\[14\text{ Guerrieri and Lorenzoni (2017), Oh and Reis (2012, and McKay and Reis (2016) were among the first contributions to this literature, focusing, respectively, on the effects of credit crunches, transfers and automatic stabilizers. Subsequent contributions include Auclert (2017), Kaplan et al. (2017), Ravn and Sterk (2017), Gornemann et al. (2016), Farhi and Werning (2016), Werning (2015) and Gali and Debortoli (2017), among many others.} \]
As discussed by Auclert (2017) that redistribution works through three channels. The *earnings heterogeneity* channel is associated with the fact that some households see their income increase more than proportionally to aggregate income, while others lose in relative terms. The *Fisher channel* refer to the fact that different households have at any point in time different net positions in nominal assets, whose real value will be affected by the change in the price level resulting from the monetary policy intervention. Finally, an additional redistributive channel arises from the likely differences across households in the mismatch between durations of assets and liabilities (including planned consumption among the latter) and, hence, on their differences in *unhedged interest rate exposure*, which thus become a third redistribution channel. As emphasized by Auclert (2017), a key determinant of the impact of each of the previous three redistribution channels is given by the size and sign of their covariance with MPCs across households, i.e. by the extent to which the redistribution caused by a monetary policy intervention favors households with a relatively high or a relatively low MPC. Those covariances can in principle be estimated using micro data on households’ consumption, income and balance sheets. The evidence reported in Auclert (2017) suggests that such redistribution channels are likely to amplify the effects of monetary policy on aggregate consumption. To see this, consider an expansionary policy which lowers real interest rates and raises output and inflation. First, the evidence suggests that households with relatively low income and wealth also tend to have relatively high MPCs. Furthermore, those households will tend to benefit more from the expansionary policy since, according to the micro evidence, (i) their earnings increase more than proportionally during output expansions, (ii) tend to have relatively large negative net nominal asset positions, and hence experience a relatively larger increase in their net wealth (in real terms) when the price level rises, and (iii) they have a lower interest rate exposure (since future asset payoffs exceed their liability counterparts, including consumption among the latter) and thus benefit more from the reduction in real interest rates. Thus, by redistributing income and wealth towards high MPC households, the three channels above work in the direction of amplifying the response of aggregate consumption to an interest rate reduction. Interestingly, as shown by Auclert (2017), the previous empirical properties can be shown to emerge, at least in qualitative terms, as an equilibrium outcome in a standard incomplete markets model à la Bewley-Hugget-Aiyagari calibrated to the U.S. economy, thus pointing to the need to introduce realistic heterogeneity in monetary models in order to capture better the effects of monetary policy.

Werning (2015) develops a general framework to identify some of the channels through which heterogeneity and incomplete markets imply a departure from the standard Euler equation for aggregate consumption. He first considers an economy with idiosyncratic risk but no borrowing or lending, and no outside assets (e.g. government debt of physical capital). In that economy, a generalized Euler equation for aggregate consumption obtains, but it is one which may imply a greater or smaller responsiveness of that variable to changes in (current and future) interest rates than its representative agent counterpart. In the particular case of household income proportional to aggregate
income (i.e. acyclical income risk) the relation between aggregate consumption and interest rates is identical to that in the RANK model. For more general specifications of the economy (e.g. with outside assets and borrowing and lending) a related “as if” result holds under specific assumptions. This is the case when liquidity is acyclical, i.e. when asset prices and/or borrowing limits move in proportion to income.

The Werning (2015) framework can be seen as a useful benchmark to understand the properties of different HANK models in the literature. Two examples, discussed by Werning himself, provide an illustration of that point. Consider the model in Ravn and Sterk (2017), which combines rigidities in price setting characteristic of NK models with search and matching in the labor market. In that framework, variations in unemployment resulting from aggregate shock generate countercyclical income risk and, as a result, an amplification of the effects of interest rate changes relative to the RANK economy. In particular, the rise in unemployment that follows an exogenous increase in interest rates leads to an increase in precautionary savings and, hence, an amplification of the effects of monetary policy relative to a RANK economy. By contrast, McKay et al. (2016) analyze a model with idiosyncratic shocks in which the effects of interest changes are dampened relative to the RANK benchmark, as a result of a built-in procyclical earnings risk (caused by the assumption of even distribution of countercyclical profits among workers) and countercyclical liquidity (resulting from constant government debt). The implied dampening of the response to monetary policy is presented by McKay et al. (2016) as a possible explanation for the forward guidance puzzle discussed in the previous section.

Kaplan et al. (2017) is another influential contribution to the literature on monetary policy and heterogeneity. A distinct feature of the HANK model developed in that paper is the coexistence of two assets: a low return liquid asset and a high return illiquid asset (e.g. housing) whose conversion into the liquid asset is subject to convex transaction costs. An implication of the latter assumption is the presence, at any point in time, of a sizable fraction of households that are wealthy but behave in a hand-to-mouth fashion since they have the bulk of their wealth in the form of illiquid assets. The presence of those households, combined with the more conventional poor hand-to-mouth, imply that a large fraction of the population is highly sensitive to labor income shocks (idiosyncratic and aggregate) but little responsive to interest rate changes. The latter fact is shown by Kaplan et al. (2017) to change dramatically the nature of the monetary policy transmission mechanism compared to the RANK model. Thus, in the HANK model the direct effect of changes in the interest rate on consumption (i.e. its effect conditional on an unchanged path for aggregate income) is much less important than its indirect effect (resulting from the induced change in aggregate income). That property is in stark contrast with the RANK model, in which the direct effect is overwhelmingly dominant, since in the latter framework all households can substitute consumption intertemporally and, under any plausible calibration, have a very small MPC.

In addition to the previous finding, the analysis in Kaplan et al. (2017) highlights an important property of HANK models: the aggregate effects of monetary policy shocks
(of, for that matter, of any other disturbance) will be shaped by the fiscal policy response to it, and, in particular, by the extent and nature of the redistributional effects of that response.

A simple approach to introducing heterogeneity and financial frictions in the NK model is provided by the so-called TANK (short for Two-Agent New Keynesian) model, in which a constant fraction of households is assumed to have no access to financial markets and just consume their current labor income, while the remaining fraction can buy and sell assets in an unconstrained way, as in the basic NK model. There are no other sources of heterogeneity within each type of households. A recent paper by Debortoli and Gali (2017) seeks to understand TANK models can provide a tractable approximation to their HANK counterparts. From a theoretical point of view, their differences and similarities are clear. They both share a key feature that is missing in representative agent models, namely, the fact that at any point in time a fraction of agents face a binding borrowing constraint (or behave as if they did). But they differ in important ways. Firstly, TANK models assume a constant fraction of constrained agents, while that fraction varies endogenously in richer HANK models. Secondly, TANK models ignore the impact on agents' current decisions of the likelihood of being financially constrained in the future, thus ruling out precautionary savings. Of course, the main advantage of TANK models relative to HANK models lies in their tractability, since there is no need to keep track of the wealth distribution and its changes over time.

Perhaps surprisingly, Debortoli and Gali (2017) show that a simple TANK model approximates well, both from a qualitative and a quantitative viewpoint, the aggregate dynamics of a canonical HANK model in response to aggregate shocks. They identify two explanations for that result. First, the TANK model approximates well the fluctuations of consumption heterogeneity between constrained and unconstrained households. Secondly, for standard calibrations of the HANK model, consumption heterogeneity within the subset of unconstrained households remains roughly constant, since those agents are able to limit consumption fluctuations by borrowing and saving.

After arguing that the TANK model constitutes a good approximation to richer and more complex HANK models, Debortoli and Gali (2017) use the former model to analyze the impact of heterogeneity on optimal monetary policy design. They show that in a simple TANK model, the objective function of a benevolent central bank can be approximated with a simple (quadratic) loss function that penalizes fluctuations in consumption heterogeneity, in addition to fluctuations in inflation and output gap. A policy trade-off emerges whenever fluctuations in (the natural level of) output are not proportionally distributed across households. In that case, it is not possible to simultaneously stabilize inflation, the output-gap and the heterogeneity index, and the central bank finds it optimal to tolerate some deviation from inflation and output from their respective targets in order to avoid too large fluctuations in consumption dispersion. For standard calibrations of the TANK model, however, the optimal policy

---

15 Early applications of the TANK framework include Gali et al. (2007) and Bilbiie (2008).
sis shown to imply minimal fluctuations in inflation and the output-gap, and is thus nearly identical to the one that would prevail in the basic NK model with a representative agent.

**Overlapping Generations**

In the previous section I have provided an overview of some recent work that, while maintaining some of the key assumptions of the NK model regarding price setting, departs from the assumption of a representative household by assuming the existence of infinite-lived households subject to idiosyncratic shocks in an environment with incomplete markets.

The assumption of an infinitely lived representative household found in the standard NK model has implications that go beyond those emphasized in the literature discussed in the previous section, which has focused on the interaction of idiosyncratic shocks and borrowing constraints to generate a non-degenerate distribution of MPCs in equilibrium, with its consequent implications for the transmission of monetary and fiscal policies. Less discussed but equally important are, in my opinion, other implications of the infinitely-lived representative household assumption.

Firstly, that assumption implies a tight link between the real interest rate and the consumer’s time discount rate along a balanced growth path. That relation all but rules out the possibility of a persistently negative natural rate of interest, with the consequent challenges that the latter would pose on a price-stability oriented monetary policy due to the zero lower bound (ZLB) on nominal interest rates. Note that in the examples from the literature on the ZLB discussed earlier the natural rate is assumed to be negative temporarily and, possibly, recurrently, but not permanently.

Secondly, the assumption of an infinitely-lived representative household rules out the existence of rational bubbles in equilibrium. The reason is simple: if a rational bubble exists in equilibrium it must grow (on average) at the rate of interest and must necessarily be held by the representative household. But if that were the case, the latter’s transversality condition would be violated, from which the impossibility of rational bubbles in that environment follows. Given the widespread view among policy makers and commentators of the important role played by the rise and fall of bubbles in economic fluctuations and, in particular, as a factor behind financial crises, and the persistent debate about how monetary policy should respond to the emergence of those bubbles, the fact that the workhorse model for monetary policy analysis cannot even account for the phenomenon of bubbles models should arguably be seen as an important shortcoming of that framework.

Three recent papers provide examples that overcome the limitations of the infinitely-lived household assumption discussed above by introducing overlapping generations of finitely-lived individuals in models with nominal rigidities. Thus, Eggertsson and

---

16 See Santos and Woodford (1997) for a proof.
Mehrotra (2014), develop a framework with three-period lived agents that makes it possible to analyze the implications for monetary policy of a (possibly permanently) negative natural rate of interest. Gali (2014, 2017) develops two alternative models with overlapping generations, monopolistic competition and sticky prices, and shows that asset price bubbles may emerge in equilibria. In both models a necessary condition for the existence of such bubbly equilibria is a natural rate of interest below the growth rate of the economy along the (unique) bubbleless balanced growth path. Most interestingly, under some conditions there may exist fluctuations in the size of the aggregate bubble which can generate aggregate fluctuations, even in the absence of shocks to fundamental. Those bubble-driven fluctuations are generally welfare-reducing. In that context one can analyze the implications of alternative monetary policy rules on fluctuations and welfare, since the evolution of bubbles is not independent of the interest rate. A central message of both papers is that a “leaning against the bubble” monetary policy, modeled as an interest rate rule that includes the size of the bubble as one of its arguments (and with a positive coefficient), is generally suboptimal and dominated by a policy that focuses on stabilizing inflation.

The Road Ahead

In the previous sections I have described a sample of recent research that extends the standard NK framework along different dimensions, in order to analyze the connection of monetary policy with a variety of phenomena (ZLB, heterogeneity, bubbles,...) that the original framework was not able to deal with.

While much of that research has been motivated, directly or indirectly, by the financial crisis of 2007-2009 and the subsequent Great Recession (and, in the case of the ZLB, by the earlier experience of Japan), there is a sense in which none of the extensions of the NK model described above can capture an important aspect of most financial crises, namely, a gradual build-up of financial imbalances leading to an eventual “crash” characterized by defaults, sudden-stops of credit flows, asset price declines, and a large contraction in aggregate demand, output and employment. By contrast, most of the models considered above share with their predecessors a focus on equilibria that take the form of stationary fluctuations driven by exogenous shocks. This is also the case in variants of those models that allow for financial frictions of different kinds and which have become quite popular as a result of the financial crisis (e.g. Bernanke et al. (1999), Christiano et al. (2014)). The introduction of financial frictions in those models often leads to an amplification of the effects of non-financial shocks. It also makes room for additional sources of fluctuations related to the presence of financial frictions (e.g. risk shocks in Christiano et al. (2014) or exogenous changes in the tightness of borrowing.

---

17 Gali (2014) assumes two-period-lived households and an inelastic labor supply (implying a constant output in equilibrium). Bubble fluctuations imply a stochastic redistribution of consumption across cohorts. By contrast, Gali (2017) introduces asset price bubbles in a perpetual youth model à la Blanchard-Yaari, conventional NK model

18 The discussion in this section draws heavily on Gali (2017).
constraints in Guerrieri and Lorenzoni (2017)). Most attempts to use a version of the NK models to explain the “financial crisis,” however, end up relying on a large exogenous shock that impinges on the economy unexpectedly, triggering a large recession, possibly amplified by a financial accelerator mechanism embedded in the model. It is not obvious what the empirical counterpart to such an exogenous shock is.

Several recent papers provide examples of model economies that are less subject to the previous criticism. A first example, albeit in the context of a real model, can be found in Boissay et al. (2016). In that paper, the authors analyze a model with asymmetric information in the interbank market, in which a sequence of small shocks may pull an economy towards a region with multiple equilibria, including equilibria characterized by a freeze in the interbank market, a credit crunch and a prolonged recession. A monetary extension of such a framework would seem highly welcome.

A second example is given by Galí (2017), which explores the possibility of fluctuations driven by stochastic bubbles in an NK model with overlapping generations. Stochastic bubbles grow at a rate above the long-term growth of the economy, generating a boom in output and employment. But they may collapse at any time with some (exogenously given) probability, pulling down aggregate demand and output. Despite its highly stylized nature, the implied equilibrium appears consistent with the pattern of asset price booms followed by sudden busts (and the induced recession) that has characterized historical financial crises. That framework, however, abstracts from any sort of financial frictions and, in particular, from the important role that high credit growth seems to have played in bubble episodes (Jordà et al. (2015)). It also leaves unexplained the factors that ultimately drive the innovations in the aggregate bubble, as well as its eventual bursting.

Basu and Bundick (2015) provide a third example of an economy with strong (and asymmetric) endogenous mechanisms behind economic fluctuations. They analyze a nonlinear version of the NK model where large and persistent slumps may arise as a result of the strong feedback between aggregate demand and (endogenous) volatility, resulting from the interaction of precautionary savings and a ZLB constraint. When the latter is absent, the central bank can adjust the nominal rate as needed in order to fully stabilize the output gap and inflation in response to demand shocks, including uncertainty shocks. That stabilization policy itself has a dampening effect on uncertainty, thus limiting precautionary savings and stimulating aggregate demand. By contrast, when the ZLB constraint is in place, there is no guarantee that the central bank will manage to stabilize the economy on the downside, which in turn raises households’ perceived volatility of future consumption (as well as its negative skewness), leading to higher precautionary savings, a reduction in output, a higher probability of falling into a liquidity trap and an additional feedback effect on volatility and skewness.

These are only examples of efforts to introduce mechanisms that may generate patterns that one may relate, at least qualitatively, to those observed in actual financial crises. I expect further research along these lines in the years ahead.
Conclusions

New Keynesian economics is alive and well. The NK model has proved to be quite flexible, with a growing number of extensions being developed by researchers in order to incorporate new assumptions or account for new phenomena. The present paper has provided a (partial) overview of some of those developments, focusing on recent work on the implications for monetary policy of (i) the ZLB constraint on the nominal interest rates, and (ii) household heterogeneity.

Despite its numerous shortcomings, most notably in terms of its ability to forecast financial crises, it is hard to think of an alternative macroeconomic paradigm that would do away with nominal rigidities and monetary non-neutralities –the two defining features of the NK model. But many challenges remain ahead, especially in finding ways of incorporating stronger endogenous propagation mechanisms that may help account for large and persistent fluctuations in output and inflation without the need to rely on large (and largely unexplained) exogenous shocks.
References


Farhi, Emmanuel and Iván Werning (2017): “Monetary Policy, Bounded Rationality and Incomplete Markets,” unpublished manuscript.


Jarocinski, Marek and Bartosz Mackowiak (2017): “Monetary Fiscal Interactions and the Euro Area’s Malaise,” unpublished manuscript.


Krugman, Paul (2009): “How did economists get it so wrong?,” *New York Times Magazine,* September 2,


Figure 1

Discretion vs Commitment in the Presence of the ZLB