The State of Macro

Olivier Blanchard

Department of Economics, MIT and NBER, Cambridge, Massachusetts 02142; email: blanchar@mit.edu

Key Words
macroeconomics, shocks, propagation mechanisms, fluctuations

Abstract
For a long while after the explosion of macroeconomics in the 1970s, the field looked like a battlefield. Over time, however, mainly because facts do not go away, a largely shared vision both of fluctuations and of methodology has emerged. Not everything is fine. Like all revolutions, this one has come with the destruction of some knowledge, and it suffers from extremism and herding. None of this is deadly, however. The state of macro is good.
1. INTRODUCTION

The editors of this new journal asked me to write about “The Future of Macroeconomics.” Nobody should accept such a task. One can forecast the near future with some confidence: Research technology is largely Austrian in nature, with output following inputs later in time. One can see the various teams at work and thus be confident that, sooner or later, they will succeed. But it is nearly impossible to forecast beyond that: Think of what somebody would have predicted for the future of macro circa 1970, before Lucas, Sargent, and Prescott came to the scene. One can, however, take stock, show evolutions, point to strengths and weaknesses of the current state of knowledge, and express hopes without disguising them as forecasts. This is what I do in this paper.

Let me continue with more caveats. I focus on only part of macro, namely fluctuations. I leave aside work on growth, where much action and much progress have taken place; this is not a value judgment, just a reflection of comparative advantage. I also make no attempt to be encyclopedic, balanced, or detached. The bibliography is largely random. In short, this is not a handbook chapter, but rather the development of a theme.

The theme is that, after the explosion (in both the positive and negative meaning of the word) of the field in the 1970s, there has been enormous progress and substantial convergence. For a while—too long a while—the field looked like a battlefield. Researchers split in different directions, mostly ignoring each other, or else engaging in bitter fights and controversies. Over time however, largely because facts have a way of not going away, a largely shared vision both of fluctuations and of methodology has emerged. Not everything is fine. Like all revolutions, this one has come with the destruction of some knowledge, and it suffers from extremism, herding, and fashion. But none of this is deadly. The state of macro is good.

The paper is organized in four sections: Section 2 sets the stage with a brief review of the past. Section 3 argues that there has been broad convergence in vision, with Section 4 looking at it in more detail. Section 5 focuses on convergence in methodology and the current challenges. One could argue that convergence in methodology is more obvious than the convergence in vision, and it should therefore have come first. This is probably true, but in the end, what matters is how we explain facts, and this is the reason for my choice of organization.

2. A BRIEF REVIEW OF THE PAST

When they launched the “rational expectations revolution,” Lucas & Sargent (1978) did not mince words:

---

1 I even leave out a topic close to my heart, the “medium run”: the low-frequency evolutions reflected in movements in capital/labor ratios, the labor share, and so on. One reason is that not much (or not enough) has happened on this front.

2 Others, I know, disagree with this optimistic assessment (for example, Solow 2008). To check whether I was totally delusional, I organized a session at the 2008 American Economic Association meetings on the theme “Convergence in Macro.” Robert Shimer (2009); Michael Woodford (2009); and V.V. Chari, together with Pat Kehoe and Ellen McGrattan (Chari et al. 2009), all wrote papers—which were published in the first issue of the new journal of the American Economic Association, the American Economic Journal: Macroeconomics and are available on the journal’s Web site. I read the papers as indeed suggesting substantial but not full convergence; the readers can judge for themselves.
That the predictions [of Keynesian economics] were wildly incorrect, and that the doctrine on which they were based was fundamentally flawed, are now simple matters of fact, involving no subtleties in economic theory. The task which faces contemporary students of the business cycle is that of sorting through the wreckage, determining what features of that remarkable intellectual event called the Keynesian Revolution can be salvaged and put to good use, and which others must be discarded.

They predicted a long process of reconstruction:

Though it is far from clear what the outcome of this process will be, it is already evident that it will necessarily involve the reopening of basic issues in monetary economics which have been viewed since the thirties as “closed” and the reevaluation of every aspect of the institutional framework within which monetary and fiscal policy is formulated in the advanced countries. This paper is an early progress report on this process of reevaluation and reconstruction.

They were right. For the next 15 years or so, the field exploded. Three groups dominated the news: the new-classicals, the new-Keynesians, and the new-growth theorists (no need to point out the PR role of “new” here), each pursuing a very different agenda.

The new-classicals embraced the Lucas-Sargent call for reconstruction. Soon, however, the Mencheviks gave way to the Bolcheviks, and the research agenda became even more extreme. Under Prescott’s leadership, nominal rigidities, imperfect information, money, and the Phillips curve, all disappeared from the basic model, and researchers focused on the stochastic properties of the Ramsey model (equivalently, a representative agent Arrow-Debreu economy), rebaptized as the Real Business Cycle (RBC) model. Three principles guided the research: explicit micro foundations, defined as utility and profit maximization; general equilibrium; and the exploration of how far one could go with no or few imperfections.

The new-Keynesians embraced reform, not revolution. United in the belief that the previous vision of macroeconomics was basically right, they accepted the need for better foundations for the various imperfections underlying that approach. The research program became one of examining, theoretically and empirically, the nature and the reality of various imperfections, from nominal rigidities, to efficiency wages, to credit market constraints. Models were partial equilibrium, or included a trivial general equilibrium closure: It seemed too soon to embody each one in a common general equilibrium structure.

The new-growth theorists simply abandoned the field (i.e., fluctuations). Lucas’s remark that, once one thinks about growth, one can hardly think about something else, convinced many to focus on determinants of growth, rather than on fluctuations and their apparently small welfare implications. Ironically, as the Ramsey growth model became the workhorse of the new-classicals, much of the progress on the growth front was made by examining the implications of various imperfections, from the public good nature of knowledge and the nature of R&D, to externalities in capital accumulation.

Relations between the three groups—or, more specifically, the first two, called by Hall “fresh water” and “salt water,” respectively (for the geographic location of most of the
new-classicals and most of the new-Keynesians)—were tense, and often unpleasant. The first accused the second of being bad economists, clinging to obsolete beliefs and discredited theories. The second accused the first of ignoring basic facts and, in their pursuit of a beautiful but irrelevant model, of falling prey to a “scientific illusion.” [See the debate between Prescott (1986) and Summers (1986).] One could reasonably despair of the future of macro [and, indeed, some of us came close (Blanchard 1992)].

This is still the view many outsiders have of the field, but it no longer corresponds to reality. Facts have a way of eventually forcing irrelevant theory out (one wishes it happened faster), and good theory also has a way of eventually forcing bad theory out. The new tools developed by the new-classicals came to dominate. The facts emphasized by the new-Keynesians forced imperfections back in the benchmark model. A largely common vision has emerged, which is the topic of the next section.

3. CONVERGENCE IN VISION

3.1. The Role of Aggregate Demand and Nominal Rigidities

It is hard to ignore facts. One major macro fact is that shifts in the aggregate demand for goods affect output substantially more than we would expect in a perfectly competitive economy. More optimistic consumers buy more goods, and the increase in demand leads to more output and more employment. Changes in the federal funds rate have major effects on real asset prices, from bond to stock prices, and, in turn, on activity.

These facts are not easy to explain within a perfectly competitive flexible-price macro model. More optimistic consumers should consume more and work less, not consume more and work more. Monetary policy should be reflected primarily in the prices of goods, not lead Wall Street to react strongly to an unexpected 25-basis-points change in the federal funds rate.

Attempts to explain these effects through exotic preferences or exotic segmented-market effects of open market operations, while maintaining the assumption of perfectly competitive markets and flexible prices, have proven unconvincing at best. This has led even the most obstinate new-classicals to explore the possibility that nominal rigidities matter. In the presence of nominal rigidities, movements in nominal money lead to movements in real money, which lead in turn to movements in the interest rate, and the demand for goods and output. With nominal rigidities, movements in aggregate demand are not automatically offset by movements in the interest rate, and thus they can translate into movements in output.

The study of nominal price and wage setting is one of the hot topics of research in macro today. It has all the elements needed to make for exciting research: It has newly available micro data sets on prices, either from CPI data bases or from large distributors (see the survey by Maćkowiak & Smets 2008). It faces delicate aggregation issues: Depending on the specific way prices are set, individual stickiness may build up or instead disappear as we look at more aggregate price indexes. The cast of characters involved in that research nicely makes the point that the old fresh water/salt water distinction has become largely irrelevant: Whereas research on the topic started with new-Keynesians, recent research has been largely triggered by an article by Golosov & Lucas (2007), itself building on earlier work on aggregation of state-dependent rules by Caplin and Caballero, among others.
3.2. Technological Shocks Versus Technological Waves

One central tenet of the new-classical approach was that the main source of fluctuations is technological shocks. The notion that there are large quarter-to-quarter aggregate technological shocks flies, however, in the face of reason. Except in times of dramatic economic transition, such as the shift from central planning to market economies in Eastern Europe in the early 1990s, technological progress is about the diffusion and implementation of new ideas and about institutional change, both of which are likely to be low-frequency movements. No amount of quarterly movement in the Solow residual will convince this skeptic: High-frequency movements in measured aggregate total factor productivity (TFP) must be due to measurement error.3

This does not imply, however, that technological progress does not play an important role in fluctuations. Though technological progress is smooth, it is certainly not constant. There are clear technological waves. Think of the high TFP growth of the post–World War II era, the low TFP growth of the 1970s and 1980s, and the higher TFP growth since the mid-1990s. These waves clearly determine movements in output in the medium and long run. But, combined with the role of anticipations on demand, and the role of demand on output, they may also determine the behavior of output in the short run. This is my next point.

3.3. Toward a General Picture and Three Broad Relations

The joint beliefs that technological progress goes through waves (i.e., that perceptions of the future affect the demand for goods today) and that, because of nominal rigidities, this demand for goods can affect output in the short run nicely combine to give a picture of fluctuations that, I believe, many macroeconomists would endorse today.

Fifty years ago, Samuelson (1955) wrote:

In recent years, 90 per cent of American economists have stopped being “Keynesian economists” or “Anti-Keynesian economists.” Instead, they have worked toward a synthesis of whatever is valuable in older economics and in modern theories of income determination. The result might be called neoclassical economics and is accepted, in its broad outlines, by all but about five per cent of extreme left-wing and right-wing writers.

I would guess we are not yet at such a corresponding stage today, but we may be getting there.

These joint beliefs are often presented in the form of three broad relations (I concentrate on a specific, more tightly specified, version—the so-called new-Keynesian model, below): (a) an aggregate demand relation, in which output is determined by demand, and demand depends in turn on anticipations of both future output and future real interest

3Too) many papers are still written with high-frequency productivity shocks as the only source of fluctuations in the model. I suspect that most of the authors use these shocks as a convenient stand-in, rather than out of conviction as to their actual existence. For some purposes, the assumption may be innocuous; if the focus is on understanding labor supply, for example, the source of shifts in labor demand—whether they be technological shocks or shifts in aggregate demand under sticky prices—may be unimportant. For other purposes, however—for example, joint movements in employment, output, and real wages—it is not and leads to artificial “puzzles,” which are puzzles only under the maintained and incorrect assumption. For a nice discussion, see Rotemberg (2008).
rates; \((b)\) a Phillips-curve like relation, in which inflation depends on both output and anticipations of future inflation; and \((c)\) a monetary policy relation, which embodies the proposition that monetary policy can be used to affect the current real interest rate (a proposition that would not hold absent nominal rigidities).

In such an economy, anticipations play a major role. Take anticipations of future productivity. The belief that a technological wave may be about to start may lead to a large increase in demand and, in turn, to a boom; think about the second half of the 1990s and the talk of a “new economy” and of an “information technology revolution.” Conversely, the realization that what looked like the start of a technological wave turns out to have been just a series of good draws may lead to a large decrease in demand and a recession [see, for example, Beaudry & Portier 2006, Lorenzoni 2009, or (I cannot resist) Blanchard et al. 2008]. Or take anticipations of inflation: The belief that inflation is well anchored limits actual movements in inflation; conversely, worries about inflation in the future may well lead to higher inflation today.

3.4. A Toy Model: The New-Keynesian Model

Within this broad picture, a specific model, the so-called new-Keynesian (NK) model, has emerged and become a workhorse for policy and welfare analysis (Clarida et al. 1999; for the application to monetary policy, see Woodford 2003).

The model starts from the RBC model without capital and, in its basic incarnation, adds only two imperfections. It introduces monopolistic competition in the goods market. The reason is clear: If the economy is going to have price setters, they better have some monopoly power. It then introduces discrete nominal price setting, using a formulation introduced by Calvo, which turns out to be the most analytically convenient. Within this frame, the three equations described earlier take a specific form:

- First, the aggregate demand equation is derived from the first-order conditions of consumers, which give consumption as a function of the real interest rate and future expected consumption. As there is no other source of demand in the basic model, consumption demand is the same as aggregate demand. Given the assumption that, so long as the marginal cost is less than the price, price setters satisfy demand at existing prices, and thus, aggregate demand is equal to output. Putting these three assumptions together, the first relation gives us output as a function of the real interest rate and future expected output.
- Second, under the Calvo specification, the Phillips curve–like equation gives inflation as a function of expected future inflation, and of the “output gap,” defined as actual output minus what output would be absent nominal rigidities.
- Third, the monetary policy rule is formalized as a Taylor rule, a reaction function giving the real interest rate chosen by the central bank as a function of inflation and the output gap. (Nominal money does not explicitly appear in the model: The assumption is that the central bank can adjust the nominal money stock so as to achieve any real interest rate it wants. What matters for activity is the real interest rate, not nominal money per se.)

The model is simple, analytically convenient, and has largely replaced the IS-LM model as the basic model of fluctuations in graduate courses (although not yet in undergraduate textbooks). Similar to the IS-LM model, it reduces a complex reality to a few simple equations. Unlike the IS-LM model, it is formally, rather than informally, derived from
optimization by firms and consumers. This has benefits and costs. The benefits are the ability to study not only activity, but also welfare, and thus to derive optimal policy based on the correct (within the model) welfare criterion. The costs are that, while tractable, the first two equations of the model are patently false (more obviously so than those in the more loosely specified IS-LM model). The aggregate demand equation ignores the existence of investment, and relies on an intertemporal substitution effect in response to the interest rate, which is hard to detect in the data on consumers. The inflation equation implies a purely forward-looking behavior of inflation, which again appears strongly at odds with the data. Still, the model yields important lessons that could not be derived in the IS-LM model and that are very general. Let me mention what I see as the main ones:

- Fluctuations in output are not necessarily bad. This was the main message of the basic RBC model, in which, indeed, all fluctuations were optimal. It remains true in the NK model. It may be best for the economy to respond to changes in technology, or changes in preferences, through some fluctuations in output and employment. Trying to smooth those fluctuations through the use of policy would be wrong.
- How relevant this argument is for rich, diversified economies remains unclear to me, and I suspect that the argument for keeping output on a smooth path is still a strong one. It is, however, surely relevant to emerging economies, which are affected by terms of trade shocks, if they are commodity exporters, or sudden shifts in capital flows. Trying to achieve a smooth path in the face of such shocks is likely, from a welfare viewpoint, to be counterproductive.
- In thinking about policy, one must think about three different concepts of activity: first, the actual level of output; second, the level of output that would prevail in the absence of nominal rigidities, often called the natural level of output (as it corresponds to the natural rate of unemployment, introduced by Friedman and Phelps), but which I prefer to call second best; and third, the level of output that would prevail in the absence of nominal rigidities and other imperfections (in the basic NK model, the monopoly power of firms). This third concept is often called the constrained efficient level of output, and, in the basic NK model, coincides with the first-best level of output.
- In the basic NK model, the monetary policy that keeps the inflation rate constant (where the price index used to measure inflation is that corresponding to the set of prices chosen by price setters) will automatically keep output at its second-best level, even in response to shocks to the supply side, such as technological shocks or oil price shocks. Is it optimal to do so? The answer depends on how the first- and second-best levels of output move in response to the shock. In the basic NK model, they move in unison. Therefore, it is indeed optimal to keep inflation constant and let output equal its second-best level.
- This leads to a strong policy conclusion: Strict inflation targeting is good, both for inflation and for output (a result Jordi Gali and I have baptized the divine coincidence). This result serves as an important benchmark. In the presence of further imperfections, however, it may no longer hold. To take a topical example, suppose labor market imperfections lead to more real wage rigidity than would be implied by a competitive labor market. Then, an increase in the price of oil—which requires a decrease in the real wage—may lead to a large decrease in the second-best level of output: Very low output and thus a large increase in unemployment may be needed to make workers accept the real wage cut. First-best output, which is defined as what output would be without real
wage rigidities, may move much less. In this case, it may be better to allow for some inflation, and a deviation of output above its second-best level for some time, rather than to stick to constant inflation.4

3.5. Building on the Toy Model

Implications of the NK model for policy, in particular monetary policy, have proven extremely rich.5 The role of anticipations in the model has allowed us to study the implications of time consistency for optimal policy, to examine the use of rules versus discretion, to discuss the role of anchoring expectations, and to think about the role of communication. Woodford’s work, including his book cited above, shows the enormous progress that has been made, and this body of work has literally changed the way central banks think about monetary policy.6

This being said, the NK model, even extended to allow, say, for the presence of investment and capital accumulation or for the presence of both discrete price and nominal wage setting, is still just a toy model, and it lacks many of the details that may be needed to understand fluctuations. The next section reviews developments on these fronts.

4. EXTENSIONS AND EXPLORATIONS

Much of the current research on macro fluctuations can be thought of an exploration of the implications of various imperfections: Beyond nominal rigidities, what are the imperfections that matter the most for macro? How do they affect the dynamic effects of shocks? How do they introduce at least the possibility of additional shocks? What do we know about these dynamic effects, and how important are these shocks? With these questions in mind, I organize this section, going market by market (from labor markets to credit and financial markets and on to goods markets), and then take up some issues that cut across markets and that I see as largely unresolved.

4.1. Labor Markets: Introducing Unemployment

One striking (and unpleasant) characteristic of the basic NK model is that there is no unemployment! Movements take place along a labor supply curve, either at the intensive margin (with workers varying hours) or at the extensive margin (with workers deciding whether or not to participate). One has a sense, however, that this may give a misleading description of fluctuations, in positive terms and, even more so, in normative terms: The welfare cost of fluctuations is often thought to fall disproportionately on the unemployed.

4 As the reader may guess from the heavier prose, I am trying to present informally some results from my own research (Blanchard & Gali 2007). My excuse is that I see these results as a good example of what can be learned from the NK model that could not be learned from the IS-LM model or its textbook extension, the AD-AS model.

5 Because the model is clearly well designed to look at monetary policy, and also perhaps because central banks are rich institutions with large research departments and conference money, there has been substantially more work on monetary policy than on fiscal policy. A good normative theory of fiscal policy in the presence of nominal rigidities remains largely to be established.

6 Another slightly cynical remark: The embrace of inflation targeting by central banks may not entirely come from their deep understanding of the new monetary theory, but instead from the coincidence of theoretical results with their long-standing desire to keep inflation low and constant.
The first question is then, How does one think about and introduce unemployment in a macro model? Here, fortunately, we can build—and are building—on a parallel effort, developed over the past 20 years by, in particular, Peter Diamond, Chris Pissarides, and Dale Mortensen [thus the name, DMP model; for a presentation, see, for example, Pissarides (2000)]. In this approach, unemployment arises from the fact that the labor market is decentralized, where, at any time, some workers are looking for jobs, while some jobs are in need of workers. This has two implications: First, by necessity, there is always some unemployment—and, symmetrically, some vacancies. Second, as it takes time for a worker to find another job, and for a firm to find another worker, both the worker and the firm have some bargaining power. This implies that the wage—and by implication, the cost of labor, employment, and unemployment—depends on the nature of bargaining.

This approach has proven extremely productive on its own. In contrast to the representative agent approach, it forces one to take into account the fact that the labor market is a market characterized by large flows, e.g., flows of job destruction and creation and flows of workers between employment, unemployment, and nonparticipation. It allows one to think about the effects of labor market institutions on the natural rate of unemployment. It also allows one to think about whether and how fluctuations affect reallocation, and whether some of the fluctuations may be due to variations in reallocation intensity. The model is sufficiently realistic in its description of the labor market that it can be confronted to the data, be it micro data on workers, micro data on firms, or, even better and increasingly available, matched panel data on workers and firms.

The central question, however, whenever we explore the implications of a specific imperfection for macro fluctuations is twofold: First, how does such an imperfection affect the dynamic effects of shocks on activity? Second, does it lead to the presence of other shocks, which may be an important source of fluctuations in activity? In the context of labor markets, we have only begun to explore the answers. Crucial to the answer is the response of real wages to labor market conditions (see, for example, Shimer 2005, Hall 2005).

Decentralized wage setting implies the existence of a wage band within which both the firm and the worker are willing to continue their relationship. The existence of such a band implies that, so long as it stays within the band, the real wage may move less than the boundaries of the band. In less formal terms, the presence of a band allows for more real wage rigidity than would be implied by a competitive labor market. This real wage rigidity does not by itself have implications for existing matches, which remain profitable so long as the wage remains within the band. If (a big if and clearly an additional assumption), however, the same real wage is also paid to new hires, then real wage rigidity has important implications for fluctuations: Combined with nominal rigidities, more real wage rigidity implies less pressure of activity on inflation; this in turn implies stronger and more persistent effects of shifts in aggregate demand, and stronger and more persistent effects of supply shocks such as increases in the price of oil, on activity.

The presence of a wage band implies that real wages can be more rigid than their competitive counterparts. The questions, however, are whether real wages, particularly the real wages of new hires, which are the wages relevant for the hiring decisions of firms, are indeed more rigid, and if so, why. This is also a hot topic of research. Theoretical work based on the exploration of constraints across workers’ wages within a firm and empirical work based on micro evidence on the wages of existing workers and new hires are proceeding apace. The next stage appears to be an integration of the market frictions that characterize the DMP model with those of efficiency wage models, which can explain
wage setting within firms and, in particular, the relation between wages paid to existing workers and to new hires.

4.2. Credit and Financial Markets

The current financial crisis makes it clear that the arbitrage approach to the determination of the term structure of interest rates and asset prices implicit in the basic NK model falls short of the mark: Financial institutions matter, and shocks to their capital or liquidity position appear to have potentially large macroeconomic effects.

The main imperfection around which thinking about credit markets is built is asymmetric information. Owners/managers of an investment project (call them the entrepreneurs) have a better knowledge of the distribution of returns on the project and of their own effort than do outside investors. As a result, outside investors are willing to participate under only certain conditions, typically when the entrepreneurs put some of their own funds into the project or put up enough collateral.

This has two direct implications for macro fluctuations. First, these constraints are likely to amplify the effects of other shocks on activity. To the extent that adverse shocks decrease profits, and thus reduce the funds available to the entrepreneurs as well as the value of the collateral they can put up, they are likely to lead to a sharper drop in investment than would happen under competitive markets. Second, shifts in the constraints can be sources of shocks. For example, changes in perceived uncertainty that lead outside investors to ask for more guarantees may prompt entrepreneurs to reduce their investment plans, leading to lower demand in the short run and lower supply in the medium run [two standard references are Bernanke & Gertler (1989), who first introduced this mechanism in an RBC model, and Kiyotaki & Moore (1997), who showed the role of asset prices and collateral].

To the extent that entrepreneurs are not financed directly by the ultimate investors but rather by financial intermediaries, who in turn get financed by the ultimate investors, these intermediaries may face the same problems as entrepreneurs. To ensure that the intermediaries have the proper incentives, the ultimate investors may want intermediaries to contribute some of their own funds. Thus, decreases in those funds will force intermediaries to decrease lending to entrepreneurs, leading again to decreases in investment (see Holmstrom & Tirole 1997). Thus, capital constraints are likely to affect both borrowers and lenders.

Furthermore, to the extent that investment projects have horizons longer than those of the ultimate investors, financial intermediaries may hold assets of a longer maturity than their liabilities. Because financial intermediaries are likely to have specific expertise about the loans they have made and the assets they hold, they may find it difficult or even impossible to sell these assets to third parties. This, in turn, opens the scope for liquidity problems: A desire by the ultimate investors to receive funds before the assets mature may force the intermediaries to sell assets at depressed prices, to cut lending, or even to go bankrupt—all possibilities the current financial crisis has made vivid. [The standard non-macro reference here is Diamond & Dybvig (1983), which has triggered a large literature.] Again, these mechanisms may amplify the effects of adverse shocks, and shifts in the distribution of funds, or in the ultimate investors’ impatience or perceptions of uncertainty, can have major macroeconomic effects.

Even in centralized markets, asset prices may not always be determined by their fundamental valuation. Many investors care about resale value, rather than just the expected
discounted value of expected payoffs on the asset. The idea that, when this is the case, infinitely long-lived assets may be subject to speculative bubbles is an old one, sustained by many apparent examples in history. A standard theoretical result is that, in the absence of other imperfections, rational speculative bubbles can exist only under dynamic inefficiency, a condition that does not appear to be satisfied in the real world. Recent research has explored whether, in the presence of other credit market imperfections, rational bubbles may exist even when the economy is dynamically efficient (Ventura 2003, Caballero et al. 2006). The conclusion is that they can.

A distinct approach to the same set of issues has explored the implications of limits to arbitrage [an argument initially formalized by Shleifer & Vishny (1997)]: Although it is easy to accept the notion that some participants in financial markets may not act rationally, the more difficult question is why others do not come to take advantage of the implied profit opportunities. The answer that has been explored in the literature is that, again because of asymmetric information between ultimate investors and potential arbitragers, the arbitragers may not have access to sufficient funds to arbitrage and return prices to fundamentals.

Yet another approach has explored the role of public and private information in the determination of asset prices and has shown that prices may respond too much to public information and too little to private information, opening the scope for large swings in prices in response to weak public signals (for example, Morris & Shin 2002, Angeletos & Werning 2006). The reason for this is that, in the presence of complementarities, investors will respond to public signals, not necessarily because they strongly believe them, but because, unlike the response to private signals, they know other investors observe them as well and may thus respond to them.

To the extent that there can be large deviations of prices from fundamentals, these can clearly be sources of shocks to activity.7 Furthermore, to the extent that these deviations are more likely to emerge in some economic environments—for example, when interest rates are low and investors are “searching for yield” or when the economy goes through a long boom and investors become steadily too optimistic (two widely held beliefs about the behavior of investors among financial market participants) such deviations will affect the dynamics of other shocks by, for example, amplifying booms and increasing subsequent slumps.

All these dimensions of credit and financial markets are also the focus of active research. Given the urgency of understanding the current financial crisis, one can be confident that progress will happen rapidly. The same is not true, however, of the next topic I discuss.

4.3. Goods Markets and Markups

In the basic NK model, the desired markup of price over marginal cost is constant. This comes from the assumption that the elasticity of substitution in utility between the differentiated goods sold by monopolistically competitive firms is constant.

Reality suggests that this assumption is wide off the mark. First, for an increasing number of goods, from software to drugs, fixed costs rather than marginal costs are the

---

7Interestingly, however, the so-called great moderation, i.e., the decline in aggregate output volatility since the early 1980s, has not been associated with a decline in aggregate stock price volatility. So far, the same is true of the current financial crisis. Although volatility has increased in financial markets, it has not yet led to increased volatility in activity.
main component of cost, with the implication that the price reflects mostly the markup rather than the marginal cost. Second, desired markups appear to be anything but constant. As neither marginal cost nor the desired markup is directly observable, the evidence here is more controversial. To my mind, perhaps the most convincing evidence comes from the findings on pass-through effects (or the lack thereof) of exchange rate movements. Recent empirical work on the United States, using disaggregated prices, shows that, when import prices are denominated in dollars at the border, exchange rate movements have minimal effect on the prices for these imports in the United States. Conversely, if import prices are denominated in foreign currency at the border, exchange rate movements lead to nearly one-for-one effects on prices for those imports (Gopinath et al. 2009). The research also shows that these differences survive long after importers have had a chance to reprice goods and thus cannot be attributed to nominal rigidities. Put another way, exchange rate movements have large and heterogeneous effects on the markups charged by importers.

How markups move, in response to what, and why, is almost terra incognita for macro.\(^{8}\) A number of theories exist. One of the most plausible may be that of consumer markets, developed and introduced in a macro model by Phelps (1994), in which firms think of the stock of consumers as an asset and choose prices accordingly. Other theories are based on games between imperfectly competitive firms. Some of these theories imply procyclical markups, so that an increase in output leads to a larger increase in the desired price, and thus to more pressure on inflation. Some imply, however, counter-cyclical markups, with the opposite implication. Some empirical evidence (for example, Bils & Chang 2000) is also available. But we are a long way from having either a clear picture or convincing theories, and this is clearly an area where research is urgently needed.

### 4.4. Some Unsettled Issues: Shocks and Anticipations

Having reviewed progress market by market, let me now slice the research another way and take up two issues I see as both central and unsettled. The first is the nature and number of major shocks behind fluctuations. The second is the actual role of anticipations. In thinking about fluctuations, an important question is whether they result from a few major sources of shocks or from many different sources, each of them with their own dynamic effects. The nature of optimal policy and the welfare implications of fluctuations depend very much on the answer.

In the traditional Keynesian interpretation of fluctuations, shocks to aggregate demand (i.e., “animal spirits”) played a major role. In the RBC interpretation, shocks to aggregate supply (i.e., technological shocks) were the major players. The quick survey of imperfections above suggests there may be many others.\(^{9}\) One can approach the question in two ways: The first approach is to use a structural model and get estimates of the shocks and of their dynamic effects. This is, for example, the approach taken by Smets & Wouters (2007), who used a state-of-the-art DSGE (dynamic stochastic general equilibrium) model.

---

\(^{8}\)I have found industrial organization economists to be reluctant to help us on this front: They seem to find the notion that one could reliably measure movements in markups over time, or the notion that one could trace and explain the evolution of an aggregate markup, both naive and doomed.

\(^{9}\)The use of “shocks” is fraught with philosophical, but also with practical, difficulties: Technological shocks, animal spirits, changes in perceived uncertainty, etc. all have deeper causes, which themselves have even deeper causes, and so on. An operational definition is that shocks are the unexplained residuals of behavioral equations. As a result, different patterns of such residuals have different effects on fluctuations.
for the U.S. economy (more on DSGE models in the next section). Their results [examined and discussed in the article by Chari et al. (2009) mentioned above] are that many shocks contribute to fluctuations, with no particular shock emerging as dominant. The problem, however, is that the answer depends very much on the specific model used to examine the data—a remark that is always true but is particularly relevant in this context. To take a trivial example, if we specify the slope of a supply curve incorrectly, we shall interpret movements along the true supply curve as deviations from our assumed supply curve, thus as shocks to supply. The tight parametrization of the DSGE models, together with the use of strong Bayesian priors, makes this risk particularly high.

The second approach is to be less structural and relies on a factor-model approach. Factor models allow one to explore whether the movements of a large number of variables can be well explained as the dynamic effects of a few underlying factors. We can then think of these factors as linear combinations of the major structural shocks. Work using factor models to interpret macro fluctuations is, I believe, just starting in earnest, and the second step, namely going from the factors to the underlying shocks [a step similar to the one going from reduced-form vector autoregressions (VARs) to structural VARs (SVARs)] remains to be taken. Nonetheless, I find some of the results we already have intriguing. Work by, for example, Stock & Watson (2005) gives the following picture: Although their formal tests find seven factors are needed to explain most of the movements in the 130 macroeconomic time series they analyze, three factors (which are orthogonal by construction) play a dominant role. The first explains most of the movements in quantities, but little of the movements in prices or asset prices. The second explains some of the movements in asset prices, but little of the movements in quantities. The third explains some of the movements in prices, but little of the movements in quantities or asset prices. The following interpretation is tempting: Shocks to aggregate demand, which move most quantities in the same direction, have little effect on prices and thus on inflation. Shocks to prices or wages, and thus to inflation, explain most of the movements in inflation, with little relation to or effect on output. Finally, asset prices largely have a life of their own, with limited effects on activity. This is not the only interpretation, nor is the work by Stock and Watson the last word, but these results do make one uneasy about the ability of monetary policy to control inflation, an assumption that underlies much of modern monetary policy.10

The second issue is the role of anticipations, which play a crucial role in the basic NK model: Other things being equal, anticipations of future consumption have a one-to-one affect on current consumption. Anticipations of future inflation affect current inflation nearly one for one as well. Under rational expectations, these imply a very large role for anticipations of future events or of future policy.

That anticipations matter a lot is obviously true. That people and firms look into the future, directly or by relying on the forecasts of others, in forming anticipations is also obviously true. Whether the basic model does not overstate the role of anticipations cum rational expectations is, however, open to question. The following are clear: Various credit

---

10The reason why this is not the only interpretation is the following: Finding in a given market that quantity and price are uncorrelated (and thus each can be explained by a separate factor, with both factors being orthogonal) is indeed consistent with a fully elastic supply curve, an inelastic demand curve, and uncorrelated demand and supply shocks (the interpretation I have implicitly given here). But it is also consistent with upward-sloping supply, downward-sloping demand curves, and the right relation between slopes and variances of the underlying supply and demand shocks.
constraints limit the ability of people and firms to spend in anticipation of good news about the future. There is a lot of adaptive learning, with people and firms looking at past evidence to update their beliefs [a route explored by Sargent (2001) and used by him to explain the evolution of the Phillips curve over time]. In addition, bounded rationality and processing abilities limit the ability of firms and households to take into account what will happen in the future [see, for example, the work by Sims (2008) and, with a slightly different formalization, an exploration of macro implications by Reis (2008)].

One reason for worry is, for example, the central role given to the anchoring of medium-term inflation expectations by central banks. The basic NK model implies that if the central bank is able to credibly anchor medium-term expectations of inflation, then the trade-off between inflation and output will be more favorable. The formal argument relies heavily on the Calvo-like specification of price setting, which implies that inflation today depends nearly one for one on inflation next year, which in turn depends on inflation in the more distant future. One may reasonably ask, however, whether price setters, choosing prices for the next month or the next quarter, will change their decision depending on what their expectation of inflation is in, say, five years. Put another way, although we very much want to believe that monetary policy can anchor inflation expectations, I am not sure we actually understand whether and how it can actually achieve it.

5. CONVERGENCE IN METHODOLOGY

That there has been convergence in vision may be controversial. That there has been convergence in methodology is not: Macroeconomic articles, whether they be about theory or facts, look very similar to each other in structure and very different from the way they did 30 years ago. The changes can be traced in part to a reaction against some of the errors of the past, but mostly to technological progress: We can solve and estimate models we could not solve before. These evolutions have been, I argue, largely but not entirely for the best.

5.1. From Small to Larger Models

Small models are essential communication and exposition devices. When successful, they reduce a complex issue to its essence. They can either embody the wisdom of larger, more explicit microfounded models, or they can trigger the development of such models. Dornbusch’s model of overshooting comes to mind as an example of the latter (Dornbusch 1976).

It is nevertheless true that much of the work in macro in the 1960s and 1970s consisted of ignoring uncertainty, reducing problems to $2 \times 2$ differential systems, and then drawing an elegant phase diagram. There was no appealing alternative—as anybody who has spent time using Cramer’s rule on $3 \times 3$ systems knows too well. Macro was largely an art, and only a few artists did it well.

Today, that technological constraint is simply gone. With the development of stochastic dynamic programming methods and the advent of software such as Dynare—a set of programs that allows one to solve and estimate nonlinear models under rational expectations (Dynare 2009)—one can specify large dynamic models and solve them nearly at the touch of a button. In many cases, larger models, even if they cannot be solved analytically, can serve the same role as the $2 \times 2$ models of lore, namely communicate a basic point or show the implications of a basic mechanism. They can deal with uncertainty
without relying on certainty equivalence, allowing us to think about such issues as precautionary saving or the behavior of investment under irreversibility. To a large extent, technological progress has reduced the required artistic component of research, and this is for the best.

5.2. From Equation-by-Equation to System Estimation

In the 1960s and early 1970s, empirical work in macro proceeded along two tracks. The first was equation-by-equation estimation of behavioral equations, be it the consumption function, the money demand equation, or the Phillips curve. The other was the development of large econometric models constructed by putting together these separate equations. Much was learned from this double-headed effort, but by the mid-1970s, the problems of these models were becoming clear. They were best identified by Sims (1980): There was little reason why the aggregate dynamics of models put together in that way would replicate actual aggregate dynamics. Putting zeros in an equation might be a good approximation for the equation as such, but not necessarily for the system as a whole.

Today, macroeconometrics is concerned mainly with system estimation, owing to the availability of new technology, namely more powerful computers. Systems, characterized by a set of structural parameters, are typically estimated as a whole. Because the likelihood function is often poorly behaved (more on this below), the standard approach is to rely on Bayesian estimation, which, in this context, can be seen as a compromise between calibration—which dominated the early RBC work—and maximum likelihood—which appears, in most cases, to ask too much of the data. VARs or SVARs—that is, VARs with a minimal set of identification restrictions that allow one to trace the effects of at least some of the structural shocks—are used in various ways: Before estimation, they are used to get a sense of the data. After estimation, they are used to compare the impulse responses to shocks implied by the structural model with those obtained from the SVAR interpretation of the data.

Because of the difficulty of finding good instruments when estimating macro relations, equation-by-equation estimation has taken a backseat—probably too much of a backseat [for estimation and discussion of the Phillips curve–like relation, see Gali et al. (2005) and the associated discussion in the corresponding issue of the *Journal of Monetary Economics*]. Another form of limited information estimation has appeared: The estimation and fitting of a subset of impulse response functions—for example the impulse response functions of various macroeconomic variables to an oil price shock—rather than estimation and fitting of the whole set of impulse response functions implied by the model. The rationale is that we may have more confidence in the estimated impulse responses to oil price shocks than in other aspects of the data and thus may prefer to use only this information to estimate the underlying parameters of the model.

5.3. DSGE Models

The most visible outcomes of this new approach are the dynamic stochastic general equilibrium (DSGE) models. They are models derived from micro foundations—that is, utility maximization by consumers-workers; value maximization by firms; rational expectations; and a full specification of imperfections, from nominal rigidities to some of the imperfections discussed above—and typically estimated by Bayesian methods. The result of estimation is a
set of structural parameters fully characterizing the model. The number of parameters has been steadily increasing with the power of computers: Smets & Wouters (2007), for example, estimate 19 structural parameters and 17 parameters corresponding to the variances and the first-order autocorrelation coefficients of the underlying shock processes.

DSGE models have become ubiquitous. Dozens of teams of researchers are involved in their construction. Nearly every central bank has one or wants to have one. They are used to evaluate policy rules, to do conditional forecasting, or, even sometimes, to do actual forecasting. There is little question that they represent an impressive achievement, but they also have obvious flaws. This may be a case in which technology has run ahead of our ability to use it, or at least our ability to use it best.

Macroeconomic data can deliver only so much information. The mapping of structural parameters to the coefficients of the reduced form of the model is highly nonlinear. Near nonidentification is frequent, with different sets of parameters yielding nearly the same value for the likelihood function—which is why pure maximum likelihood is rarely used [on nonidentification or weak identification, see Canova & Sala (2006) or Iskrev (2008)]. The use of additional information, as embodied in Bayesian priors, is clearly conceptually the right approach. However, in practice, the approach has become rather formulaic and hypocritical. The priors used often reflect the priors of others and, after backward recursion has traced their origins, have little basis in facts. Partly for the same reason, models are also more similar in their structure than would seem desirable: Roughly the same models are used both in rich and in emerging economies, despite their different structures and shocks.

Current theory can also deliver only so much. One of the principles underlying DSGE models is that, in contrast to the previous generation of models, all dynamics must be derived from first principles.11 The main motivation is that, only under these conditions, can welfare analysis be performed. A general characteristic of the data, however, is that the adjustment of quantities to shocks appears slower than implied by our standard benchmark models. Reconciling the theory with the data has led to a lot of unconvincing reverse engineering. External habit formation—that is, a specification of utility where utility depends not on consumption but on consumption relative to lagged aggregate consumption—has been introduced to explain the slow adjustment of consumption. Convex costs of changing investment, rather than the more standard and more plausible convex costs of investment, have been introduced to explain the rich dynamics of investment. Backward indexation of prices, an assumption that, as far as I know, is simply factually wrong, has been introduced to explain the dynamics of inflation. Because their introduction can then be blamed on others, these assumptions have often become standard, passed on from model to model with little discussion. This way of proceeding is clearly wrongheaded.

First, such additional assumptions should be introduced in a model only if they have independent empirical support. The fact that an additional assumption helps fit the aggregate dynamics in a model that is surely misspecified elsewhere is not convincing.

Second, it is clear that heterogeneity and aggregation can lead to aggregate dynamics that have little apparent relation to individual dynamics [see Chang et al. (2008) on the relation of the aggregate labor supply relation to individual labor supply, when individual labor supply decisions are taken both at the intensive and the extensive margin and work-

---

11This is not as clean a position as it sounds, as shocks are typically allowed to have their own, unexplained, dynamics, for example, to follow AR(1) processes.
ers have limited access to insurance]. Progress is being made, both theoretically and empirically, in deriving aggregate dynamic implications from individual behavior [for a recent review and exploration, see Caballero & Engel (2007)]. Until further progress is made, it may well make more sense to recognize our ignorance and to allow part of the dynamics of our DSGE models to be data determined. True, this would make formal welfare analysis impossible, but welfare analysis based on the wrong interpretation of the data is clearly worse. For example, it matters very much for the assessment of the welfare costs of fluctuations whether the slow adjustment of consumption is attributed to habit formation, aggregation, or slowly adjusting expectations. Ad-hoc welfare functions in terms of deviations of inflation and deviations of output from some smooth path may be the best we can do given what we know.

6. IN GUISE OF A CONCLUSION

I have argued that macroeconomics is going through a period of great progress and excitement and that there has been convergence in both vision and methodology over the past two decades. There is, however, such a thing as too much convergence. To caricature, but only slightly: A macroeconomic article today often follows strict, haiku-like rules. It starts from a general equilibrium structure, in which individuals maximize the expected present value of utility, firms maximize their value, and markets clear. Then, it introduces a twist, be it an imperfection or the closing of a particular set of markets, and works out the general equilibrium implications. It then performs a numerical simulation based on calibration, showing that the model performs well. It ends with a welfare assessment.

Such articles can be great, and the best ones indeed are. More often than not, however, they suffer from some of the flaws I discussed above in the context of DSGE models: introduction of an additional ingredient in a benchmark model already loaded with questionable assumptions and little or no independent validation for the added ingredient. Thus, I end this review with three related hopes/pleas.\(^\text{12}\)

The first is for the rehabilitation of partial equilibrium modeling in macroeconomics. Although it is only a first step, it is important to understand the implications of a particular imperfection on its own, i.e., taking as given a large part of the macroeconomic environment. Forcing oneself to examine the implications of this imperfection in general equilibrium from the start typically creates unattractive trade-offs. For example, many imperfections lead to heterogeneity of income and wealth across agents; a general equilibrium closure requires the introduction of various auxiliary assumptions, such as counterfactual assumptions about the existence of various forms of insurance, making it difficult to assess the relative roles of the central and the auxiliary assumptions. Better in this case to proceed in two steps, with partial equilibrium first, and with the—admittedly much tougher—general equilibrium problem second.\(^\text{13}\)

The second is that no additional ingredient should be introduced in a general equilibrium model without some independent validation. We have increased access to large microdata sets, which allow us to learn about aspects of individual behavior. We are steadily

\(^{12}\)Andrei Shleifer has pointed out to me that my three pleas are, to borrow an expression from computer design, for a more “open architecture” of the field.

\(^{13}\)To quote Solow (2008): “My general preference is for small, transparent, tailored models, often partial equilibrium, usually aimed at understanding some little piece of the (macro-)economic mechanism.”
deriving theories of aggregation, which allow us to derive the dynamic aggregate implications of individual behavior [for an example in the case of investment, see Caballero et al. (1995)]. To the extent possible, this should be how we proceed. Introducing more realistic descriptions of aggregate behavior in a DSGE model should be a last step, not a first step.

The third is for the relegalization of shortcuts and simple models. DSGE models tend to be very complex. Approximating complex relations by simple ones helps intuition and communication. The shortcuts of the past may have been potentially dangerous, to be used only by the masters of the trade, but the job is now potentially much easier. We can start from fully articulated models and see whether we can capture their essence through simpler relations. We can then check whether and when the implications of the shortcut fit the main characteristics of the full-fledged model [a nice example in this context is provided by Krusell & Smith (1998)]. We should be willing to do more of this than we are today.

DISCLOSURE STATEMENT

The author is not aware of any biases that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

Thanks to Daron Acemoglu (committee member for the Annual Review of Economics), Ricardo Caballero, V.V. Chari, Laura Feiveson, Ben Friedman, Jordi Gali, Anil Kashyap, Jaewoo Lee, N. Greg Mankiw, Marianna Riggi, Julio Rotemberg, Robert Shimer, Andrei Shleifer, and Harald Uhlig for their very useful comments.

LITERATURE CITED


Caballero R, Engel E. 2007. Price stickiness in Ss models; new interpretations of old results. J. Monet. Econ. 54 (Suppl. 1):100–21


