

WHAT DO WE KNOW ABOUT MACROECONOMICS THAT FISHER AND
WICKSELL DID NOT?***

BY

OLIVIER BLANCHARD*

Summary

The answer to the question in the title is: A lot. In this essay, I argue that the history of macroeconomics during the 20th century can be divided into three epochs:

Pre-1940. A period of exploration, where macroeconomics was not macroeconomics yet, but monetary theory on one side, business cycle theory on the other. A period during which all the right ingredients, and quite a few more, were developed. But also a period where confusion reigned, because of the lack of an integrated framework.

From 1940 to 1980. A period of consolidation. A period during which an integrated framework was developed – starting with the *IS-LM*, all the way to dynamic general equilibrium models – and used to clarify the role of shocks and propagation mechanisms in fluctuations. But a construction with an Achilles's heel, namely too casual a treatment of imperfections, leading to a crisis in the late 1970s.

Since 1980. A new period of exploration, focused on the role of imperfections in macroeconomics, from the relevance of nominal price setting, to incompleteness of markets, to asymmetric information, to search and bargaining in decentralized markets. Exploration often feels like confusion. But behind it may be one of the most productive periods of research in macroeconomics.

Key words: macroeconomics, business cycles, fluctuations, history of thought

1 INTRODUCTION

The theme of this lecture was triggered by a question from the editors of the *Quarterly Journal of Economics*. To celebrate the new century, they commissioned a number of essays on the theme: What do we know about field x that Marshall did not know?¹ I took up the challenge for macroeconomics and this lecture is the result.

* MIT and NBER.

*** Thirteenth Tinbergen lecture, delivered on October 15, 1999 at De Nederlandsche Bank, Amsterdam. I thank Daron Acemoglu, Ben Bernanke, Ricardo Caballero, Thomas Cool, Peter Diamond, Rudi Dornbusch, Stanley Fischer, Bengt Holmström, Larry Katz, David Laibson, N. Greg Mankiw, David Romer, Paul Samuelson, Andrei Shleifer, Robert Solow, Justin Wolfers, and Michael Woodford for useful comments and discussions. This lecture is a slightly modified version of an article with the same title, forthcoming in the *Quarterly Journal of Economics*, November 2000.

1 The essays have been and will be published in the various issues of the journal during 2000.

The first thing to say is that, in the case of macroeconomics and with due respect to Marshall, he is probably not the right reference.² But if we replace his name by those of Wicksell and Fisher, the two dominant figures in the field at the start of the 20th century, then the answer is very clear: We have learned a lot. Indeed, progress in macroeconomics may well be the success story of 20th century economics.³

Such a strong statement will come as a surprise to some. On the surface, the history of macroeconomics in the 20th century appears as a series of battles, revolutions and counterrevolutions, from the Keynesian revolution of the 1930s and 1940s, to the battles between Monetarists and Keynesians of the 1950s and 1960s, to the Rational Expectations revolution of the 1970s, and the battles between New Keynesians and New Classicals of the 1980s. These suggest a field starting anew every twenty years or so, often under the pressure of events, and with little or no common core. But this would be the wrong image. The right one is of a surprisingly steady accumulation of knowledge. The most outrageous claims of revolutionaries make the news, but are eventually discarded. Some of the others get bastardized, then integrated. The insights become part of the core. In this article, I focus on the accumulation of knowledge rather than on the revolutions and counterrevolutions. Admittedly, this makes for worse history of thought, and it surely makes for worse theater. But it is the best way to answer the question in the title.⁴

Let me state the thesis that underlies this essay. I believe the history of macroeconomics during the 20th century can be divided in three epochs, the third one currently playing:⁵

- Pre 1940. A period of exploration, where macroeconomics was not macroeconomics yet, but monetary theory on one side, business cycle theory on the other. A period during which all the right ingredients, and quite a few more, were developed. But also a period where confusion reigned, because of the lack of an integrated framework.

2 Marshall published two books about macroeconomics, the first early in his professional career, *The Economics of Industry*, in 1879; the second one year before his death in 1924, *Money, Credit and Commerce*, which largely repeated the themes of the first. In his 'in memoriam' biography of Marshall, Keynes argues that Marshall anticipated many of the points developed by Fisher and Wicksell. But his assessment surely reflects in part the kindness required under the circumstances.

3 The two most relevant books here are Wicksell's *Interest and Prices*, published in 1898 in German, but translated in English only much later, and Fisher's *Purchasing Power of Money*, published in 1911.

4 A nice, largely parallel, review of macroeconomics in the 20th century, taking the alternative, more historical, approach is given by Woodford (1999).

5 For the purpose of this article, I shall define macroeconomics as the study of fluctuations, mundane – recessions and expansions, or sustained – sharp recessions, long depressions, sustained high unemployment. I shall exclude both the study of growth and the political economy of macroeconomics. Much progress has been made there as well, but covering these two topics would extend the length of this essay to unmanageable proportions.

- From 1940 to 1980. A period of consolidation. A period during which an integrated framework was developed – starting with the *IS-LM*, all the way to dynamic general equilibrium models – and used to clarify the role of shocks and propagation mechanisms in fluctuations. But a construction with an Achilles’s heel, namely too casual a treatment of imperfections, leading to a crisis in the late 1970s.
- Since 1980. A new period of exploration, focused on the role of imperfections in macroeconomics, from the relevance of nominal wage and price setting, to incompleteness of markets, to asymmetric information, to search and bargaining in decentralized markets, to increasing returns in production. Exploration often feels like confusion, and confusion there indeed is. But behind it may be one of the most productive periods of research in macroeconomics.

Let me develop these themes in turn.

2 PRE-1940. EXPLORATION

To somebody who reads it today, the pre-1940 literature on macroeconomics feels like an (intellectual) witch’s brew: many ingredients, some of them exotic, many insights, but also a great deal of confusion.⁶

The set of issues that would now be called macroeconomics fell under two largely disconnected headings: Monetary Theory, and Business Cycle Theory.^{7, 8}

At the center of Monetary Theory was the quantity theory – the theory of how changes in money lead to movements in output and prices. The focus was both on long-run neutrality and on short-run non-neutrality. The discussion of the short-run effects of an increase in money on output was not much improved relative to, say, the earlier treatments by Hume or Thornton. Some stressed the effects from money to prices and from prices to output: Higher money led to higher

6 This assessment should be taken with a grain of salt. Paul Samuelson once chided Stanley Fischer and me for having, in our graduate macro text, few if any references to articles written before 1980. Some of the confusion may therefore be in the eyes of this ignorant beholder.

7 The word ‘macroeconomics’ does not appear until the 1940s. According to JSTOR (the electronic data base which includes the articles from most major journals since their inception), the first use of ‘macro-economic’ in the title of an article is by De Wolff in 1941, in ‘Income Elasticity of Demand, a Micro-economic and a Macro-economic Interpretation’; the first use of ‘macroeconomics’ in the title of an article is by Klein in 1946, in an article called, fittingly, ‘Macroeconomics and the Theory of Rational Behavior.’ But the word ‘macrodynamics’ can be found in writings by Frisch and Kalecki in the early 1930s. And Paul Samuelson tells me that, in an analogy with thermodynamics, his Harvard teacher, Gottfried Haberler was using the words ‘macroscopic’ and ‘microscopic’ to indicate different ways of looking at the economy.

8 In the 1920s, the two graduate ‘macro’ courses at Harvard University were called ‘Principles of Money and Banking’ and ‘Commercial Crises.’ The latter course became ‘Economic Trends and Fluctuations’ in 1930, and ‘Business Cycles and Economic Forecasting’ in 1934. Except for 1933, the year in which Schumpeter taught a unified course, this separation held until the 1940s.

prices, higher prices ‘excited’ business and led in turn to higher output. Others stressed the effects from money to output, and from output to prices: Higher money increased demand and output, and the increase in output led in turn to an increase in prices over time.

Business Cycle Theory was not a theory at all, but rather a collection of explanations, each with its own rich dynamics.⁹ Most explanations focused on one factor at a time: Real factors (weather, technological innovations), or expectations (optimistic or pessimistic firms), or money (banks or the central bank). When favorable, these factors led firms to invest more, banks to lend more, until things turned around, typically for endogenous reasons, and the boom turned into a slump. Even when cast as general equilibrium, the arguments, when read today, feel incomplete and partial equilibrium in nature: It is never clear how, and in which markets, output and the interest rate are determined.

In retrospect, one can see the pieces of a macroeconomic model slowly falling into place.¹⁰

At the center was the difference, emphasized by Wicksell, between the natural rate of interest (the rate of return on capital) and the money rate of interest (the interest rate on bonds). This would become a crucial key in allowing for the eventual integration of goods markets (where the natural rate is determined) and financial markets (where the money rate is determined). It would also prove to be the key in allowing for the eventual integration of monetary theory (where an increase in money decreases the money rate relative to the natural rate, triggering higher investment and higher output for some time) and business cycle theory (in which several factors, including money, affect either the natural rate or the money rate, and thus the difference between the two).

Where the literature remained confused, at least until Keynes and for some time after, was how this difference between the two rates translated into movements in output. Throughout the 1920s and 1930s, the focus was increasingly on the role of the equality of saving and investment, but the semantic squabbles which dominated much of the debate (the distinctions between ‘*ex ante*,’ and ‘*ex post*,’ ‘planned’ and ‘realized’ saving and investment, the discussion of whether the equality of saving and investment was an identity or an equilibrium condition) reflected a deeper confusion. It was just not clear how shifts in saving and investment affected output.

In that context, the methodological contributions of the *General Theory* (1936) made a crucial difference:

9 The variety and the complexity of these explanations is reflected in Mitchell (1923), or in the textbook of the time, *Prosperity and Depression* by Haberler (1937).

10 There is always the risk here of seeing more than there actually was. An example of this is Alvin Hansen’s textbook, published in the early 1950s, of the history of macroeconomic thought through a thoroughly Keynesian filter: How close various authors had gotten to $S = I$, to what determined S , to what affected I . There is little question that the filter is highly distorting.

- Keynes explicitly thought in terms of three markets (the goods, the financial, and the labor markets), and of the implications of equilibrium in each.
- Using the goods market equilibrium condition, he showed how shifts in saving and investment led to movements in output.
- Using equilibrium conditions in both the goods and the financial markets, he then showed how various factors affected the natural rate of interest (which he called the ‘marginal efficiency of capital’), the money rate of interest, and output. An increase in the marginal efficiency of capital – coming, say, from more optimistic expectations about the future – or a decrease in the money rate – coming from expansionary monetary policy – both led to an increase in output.

A quote from Pigou’s Marshall lectures, ‘Keynes’s *General Theory*, A Retrospective View,’ (1950) puts it well:¹¹

‘Nobody before him, so far as I know, had brought all the relevant factors, real and monetary at once, together in a single formal scheme, through which their interplay could be coherently investigated.’

The stage was then set for the second epoch of macroeconomics, a phase of consolidation and enormous progress.

3 1940-1980. CONSOLIDATION.

Macroeconomists often refer to the period from the mid 1940s to the mid 1970s as the golden age of macroeconomics. For a good reason: Progress was fast, and visible.

3.1 *Establishing a Basic Framework*

The *IS-LM* formalization by Hicks and Hansen may not have captured exactly what Keynes had in mind. But, by defining a list of aggregate markets, writing demand and supply equations for each one, and solving for the general equilibrium, it transformed what was now becoming ‘macroeconomics.’ It did not do this alone. Equally impressive in their powerful simplicity were, for example, the model developed by Modigliani in 1944, with its treatment of the labor market and the role of nominal wage or price rigidities, or the model developed by Metzler in 1951, with its treatment of expectations, wealth effects, and the government budget constraint. These contributions shared a common structure: The reduction of the economy to three sets of markets – goods, financial, and labor – and a focus on the simultaneous determination of output, the interest rate and the

¹¹ Pigou’s first assessment of the *General Theory*, in 1936, had been far less positive, and for understandable reasons: Keynes was not kind to Pigou in the *General Theory*; but, by 1950, time had passed, and Pigou clearly felt more generous.

price level. This systematic, general equilibrium approach to the characterization of macroeconomic equilibrium became the standard, and, reading the literature, one is struck at how much clearer discussions became once this framework had been put in place.

This approach was brought to a new level of rigor in *Money, Interest, and Prices* by Patinkin (1956). Patinkin painstakingly derived demand and supply relations from intertemporal optimizing behavior by people and firms, he characterized the equilibrium, and, in the process, laid to rest many of the conceptual confusions which had plagued earlier discussions. It is worth making a – non limitative – list (if only because some of these confusions have a way of coming back in new forms):

- ‘Say’s law’: False. In the same way as the supply of any particular good did not automatically generate its own demand (the relative price of the good has to be right), the supply for all goods taken together did not generate its own demand either: The intertemporal price of goods, the real interest rate, also had to be right.
- ‘Walras’ law’: True. So long as each agent took all his or her decisions under one budget constraint, then equilibrium in all markets except one implied equilibrium in the remaining one.
- The ‘Classical Dichotomy’ between the determination of the price level on the one hand, and the determination of relative prices on the other: False. ‘Neutrality,’ the proposition that changes in money were ultimately reflected in proportional changes in the price level, leaving relative prices unaffected, was true. But this was an equilibrium outcome, not the result of a dichotomous model structure.
- ‘Value theory *versus* Monetary Theory’ – the issue of whether standard methods used in value theory could be used to think about the role and the effects of money in a monetary economy. The answer was: Yes. One could think of real money balances as entering either the indirect utility of consumers, or the production function of firms. One could then treat real money balances as one would treat any other good.
- ‘Loanable Funds or Liquidity Preference’ – the issue of whether the interest rate was determined in the goods markets (through the equality of saving and investment), or in the financial markets (through the equality of the demand and the supply of money). The answer, made clear by the general equilibrium structure of the models, was: In general, both.

3.2 *Back to Dynamics*

Keynes himself had focused mostly on comparative statics. Soon after however, the focus shifted back to dynamics. Little if any of the old business cycle literature was used, and most of the work was done from scratch.

Key to these developments was the notion of ‘temporary equilibrium,’ developed by Hicks in *Value and Capital* (1939). The approach was to think of the economy as an economy with few future or contingent markets, an economy in which people and firms therefore had to make decisions based partly on state variables – variables reflecting past decisions – and partly on expectations of the future. Once current equilibrium conditions were imposed, the current equilibrium depended partly on history, partly on expectations of the future. And given a mechanism for the formation of expectations, one could trace the evolution of the equilibrium through time.

Within this framework, the next step was to look more closely at consumption, investment, and financial decisions, and their dependence on expectations. This was accomplished, in a series of extraordinary contributions, by Modigliani and Friedman who examined the implications of intertemporal utility maximization for consumption and saving, by Jorgenson and Tobin who examined the implications of value maximization for investment, and by Tobin and a few others who examined the implications of expected utility maximization for financial decisions. These developments would warrant more space, but they are so well known and recognized (in particular, by many Nobel prizes) that there is no need to do so here.

The natural next step was to introduce rational expectations. The logic for taking that step was clear: If one was to explore the implications of rational behavior, it seemed reasonable to assume that this extended to the formation of expectations.¹² That step however took much longer. It is hard to tell how much of the delay was due to technical problems – which indeed were substantial – and how much to objections to the assumption itself. But this was eventually done, and by the late 1970s, most of the models had been reworked under the assumption of rational expectations.¹³

With the focus on expectations, a new battery of small models emerged, with more of a focus on intertemporal decisions. The central model was a remake of a model first developed by Ramsey in 1928, but now reinterpreted as a temporary equilibrium model with infinitely lived individuals facing a static production technology.¹⁴ This initial structure was then extended in many directions.¹⁵ Among these:

12 As Solow (1999) put it: ‘The big intellectual advantage of rational expectations is that it is definite. The complement of rational expectations is a very large grab-bag.’

13 This is where a more historical approach would emphasize that this was not a smooth evolution... At the time, the introduction of rational expectations was perceived as an attack on the received body of macroeconomics. But, with the benefit of hindsight, it feels much less like a revolution than like a natural evolution. (Some of the other issues raised by the same economists who introduced rational expectations proved more destructive, and are at the source of the crisis I discuss below.)

14 Ramsey had thought of his model as purely normative, indicating how a central planner might want to allocate consumption over time.

- The introduction of costs of adjustment for capital, leading to a well defined investment function, and a way of thinking about the role of the term structure of interest rates in achieving the equality of saving and investment.
- The introduction of money as a medium of exchange, and the extension of the Baumol-Tobin model of money demand to general equilibrium.
- The introduction of some dimensions of heterogeneity, for example allowing for finite lives and extending the overlapping generations model first developed by Samuelson and Diamond.
- The introduction of a leisure/labor choice, in addition to the consumption/saving decision.
- The extension to an economy open both in goods and financial markets.

Initially, these models were solved either under perfect foresight, a simplifying but rather unappealing assumption in a world of uncertainty and changing information. That introducing uncertainty was essential was driven home by an article by Hall (1978) who showed that, under certain conditions, optimizing behavior implied that consumption should follow a random walk – a result which initially came as a shock to those trained to think in terms of the life cycle model. Under the leadership of Lucas and Sargent (for example, Lucas and Stokey (1989), Lucas (1987), and Sargent (1987)), developments in stochastic dynamic programming together with progress in numerical methods and the development of more powerful computers were used to characterize behavior under uncertainty. This in turn allowed the exploration of a new and important set of issues, the implications of the absence of some future or contingent markets in affecting consumption and investment decisions, and, in turn, the macroeconomic equilibrium.

Compared to the first generation of models (the *IS-LM*, Metzler, and Modigliani models), these models, in either their perfect foresight or their stochastic versions, were more tightly specified, less eclectic. In their initial incarnation, they often ignored imperfections which many macroeconomists saw as central to an explanation of macroeconomic fluctuations. But they provided a basic set of off-the-shelf structures on which to build, and indeed, in which to introduce imperfections. Getting ahead of my story: This is indeed what has happened since the early 1980s, and I shall return to it later.

3.3 *From Models to Data*

Starting in the 1940s, in macroeconomics as in the rest of economics, research was radically transformed by the increasing availability of data and the development of econometric methods. But another element, specific to macroeconomics,

15 The basic Ramsey model and many of these extensions form the core of today's graduate textbooks. See for example Blanchard and Fischer (1989), or Obstfeld and Rogoff (1996).

played a central role: the coincidence between the implications of linear dynamic models and the time series representation of economic variables.

The old business cycle literature had been groping towards non-linear endogenous cycle models, models where the expansion created the conditions for the next recession, the recession the conditions for the next expansion, and so on. As early as 1933 however, Ragnar Frisch had argued that much simpler systems, linear difference systems with shocks, could provide a better account of aggregate fluctuations. In an economy described by such systems, one could think of fluctuations as the result of the combination of impulses – random shocks constantly buffeting the economy – and propagation mechanisms, the dynamic effects of these shocks implied by the linear system. This point was reinforced by Samuelson's 1939 analysis of the multiplier accelerator, which showed how a given shock to spending could generate rich dynamic responses of output. The convenience of the approach, and its easy mapping to the data, quickly led to the dominance of linear models with shocks as the basic approach to fluctuations, and alternative non-linear approaches largely faded from the scene.

These early steps were followed by the specification and estimation of structural models. Using the approach to identification developed by Koopmans and others at the Cowles Commission, individual equations for consumption, investment, and money demand, were estimated and integrated in larger and larger macroeconomic models, starting with the work of Tinbergen in the 1930s, and culminating, on the academic side, in models such as the MPS, developed by Modigliani and co-authors in the 1960s and early 1970s.

In the late 1970s, the focus shifted, at least on the academic side, to smaller models. The feeling was that the immense effort to construct large structural models had been overambitious, that the identification conditions used in estimation of individual equations were often dubious, and that the equation-by-equation construction of econometric models did not in any way insure that the reduced form of the estimated model fitted the basic characteristics of the data.

This was the motivation behind the return to smaller, more transparent, structural models, whose limited size had the additional advantage of making them solvable under rational expectations. It was also the motivation behind the development of a new statistical tool, vector autoregressions or VARs – namely the direct estimation of the joint stochastic process describing the variables under consideration. VARs were then used in two ways: to obtain a set of stylized statistical facts that models had to match and to see whether, under a minimal set of identification restrictions, the evidence was consistent with the dynamic effects of shocks implied by a particular theory or class of theories.

The constant back and forth between models and data, and the increasing availability of macro and micro data, has made macroeconomics a radically different field from what it was in 1940. Samuelson once remarked that one of his disappointments was that econometric evidence had led to less convergence than he had hoped when the first econometric steps were taken (in Snowdown and Vane

(1999), p 323). It is nevertheless true that progress has been nothing short of amazing. When Kahn (1931) first tried to get a sense of the value of the marginal propensity to consume, all he had were a few observations on proxies for aggregate production, imports, and investment. When Modigliani and Brumberg (1954) tried to assess the empirical fit of the life cycle hypothesis, they could use the time series recently put together for the National Income Accounts by Kuznets and others, and a few cross sections on income and saving. Today, studies of consumption have access to long repeated cross sections, or even long panel data sets (see for example Deaton (1992)). This allows not only for much sharper questions about consumption behavior, but also for a more convincing treatment of identification (through the use of ‘natural experiments’, tracing the effects of changes in the economic environment affecting some but not all consumers) than was feasible earlier.

So far, the tone of this essay has been that of a panegyric, the description of a triumphal march towards truth and wisdom. Let me now turn to the problem that macroeconomics largely ignored, and which led to a major crisis in the late 1970s.

3.4 *The Casual Treatment of Imperfections*

From Keynes on, there was wide agreement that some imperfections played an essential role in fluctuations.¹⁶ Nominal rigidities, along the lines suggested by Keynes, and later formalized by Modigliani and others, played an explicit and central role in most formalizations. They were crucial to explaining why and how changes in money and other shifts in the demand for goods affected output, at least in the short run.

These nominal rigidities, when combined with later developments such as rational expectations, proved to have rich and relevant implications. For example, in an extension of the Mundell-Fleming model, (the version of the *IS-LM* model for an economy open in both goods and financial markets) Dornbusch (1976) showed that the large swings in exchange rates, which had been observed after the adoption of flexible exchange rates in the early 1970s and were typically attributed to irrational speculation, could be interpreted instead as the result of arbitrage by speculators with rational expectations in an economy with a slowly adjusting price level. The lesson was more general: nominal rigidities in some markets led to more price volatility in others, here in the foreign exchange market.

16 A semantic clarification: Following tradition, I shall refer to ‘imperfections’ as deviations from the standard perfect competition model. Admittedly, there is more than just a semantic convention here. Why give such status to such an utterly unrealistic model? The answer is because most current research is organized in terms of what happens when one relaxes one or more assumptions in that model. This may change one day. But, for the time being, this approach provides a common research strategy and makes for easier communication among macroeconomic researchers.

But, as the early models were improved in many dimensions, the treatment of imperfections remained surprisingly casual. The most obvious example was the treatment of wage adjustment in the labor market. In early models, the assumption was typically that the nominal wage was fixed, and that the demand for labor then determined the outcome. Later on, these assumptions were replaced by a Phillips curve specification, linking inflation to unemployment. But there was surprisingly little work on what exactly laid behind the Phillips curve, why and how wages were set this way, why there was little apparent relation between real wages and the level of employment. As a result, most macro models developed in the 1960s and 1970s had a schizophrenic feeling: a careful modeling of consumption, investment, and asset demand decisions on the one hand, and a theoretical specification of price and wage setting on the other.

The only systematic theoretical attempt to explore the implications of nominal rigidities, the ‘fixed price equilibrium’ approach developed in the 1970s, turned out to be a dead end. This was for reasons intrinsic to the macroeconomic approach at the time, and so it is worth looking at the episode more closely.

The approach was based on the insights of Clower and Patinkin, and a systematic treatment was given by Barro and Grossman (1976). The strategy was to assume a competitive economy, to allow the vector of prices to differ from the flexible price equilibrium vector, and then to characterize the determination of output under these conditions. Equilibrium in each market was assumed equal to the minimum of supply and demand. The complexity and the richness of the analysis came from the fact that, if people or firms were on the short side in one market, they then modified their supply and demand functions in other markets.

The results were tantalizing. The analysis showed that, if the price vector was such as to yield a state of generalized excess supply (in both goods and labor markets), the economy behaved very much as in the Keynesian model. Firms, constrained in the goods market, employed only as many workers as they needed: The demand for labor was determined by output, and was independent of the real wage. Workers, constrained in the labor market, took employment and labor income as given in making consumption decisions. The economy exhibited a demand multiplier: Increases in demand led to more production, more income, more demand, and so on.

This was clearly good news for the prevailing view of fluctuations. But the same approach showed that, if the price vector was such as to yield instead a state of generalized excess demand, things looked very different, indeed much more like what one saw in the Soviet Union at the time than in market economies. The multiplier was a supply multiplier: The inability to buy goods led people to reduce their labor supply, decreasing output, leading to even more rationing in the goods market, and so on. An increase in government spending led to more rationing of consumers, a decrease in labor supply, and a decrease in output.

This raised an obvious issue: Without a theory of why prices were not right in the first place, the second outcome appeared just as likely as the first. So, why was it that we typically observed the first outcome, not the second? The answer clearly required a theory of price setting, and so the explicit introduction of price setters (so that somebody other than the auctioneer was in charge). But if there were explicit price setters, there was then no particular reason why the market outcome should be equal to the minimum of supply and demand. If, for example, price setters were monopolistic firms, and demand turned out larger than they expected, then they might well want to satisfy this higher level of demand, at least as long as their price exceeded marginal cost. So, to make progress, one had to think hard about market structure, and who the price setters were. But such focus on market structure, and on imperfections more generally, was altogether absent from macroeconomics at the time.¹⁷

At roughly the same time (circa 1975), this intellectual crisis was made worse by another development, the collapse of traditional conclusions when rational expectations were introduced in otherwise standard Keynesian models. Working within the standard model at the time (an *IS-LM* model plus an expectations-augmented Phillips curve), Sargent (1973) showed that, if one assumed rational expectations of inflation, the effects of money on output lasted only for a brief moment, until the relevant information about money was released. So, even on its own terms, once rational expectations were introduced, the standard model seemed unable to deliver its traditional conclusions (such as for example, lasting effects of money on output).

Thus, by the end of the 1970s, macroeconomics faced a serious crisis. The reaction of researchers was to follow two initially very different routes.

The first, followed by the 'New Keynesians', was based on the belief that the traditional conclusions were indeed largely right, and that what was needed was a deeper look at imperfections and their implications for macroeconomics.

The second, followed by the 'New Classical' or 'Real Business Cycle' theorists, was instead to question the traditional conclusions and explore how far one could go in explaining fluctuations without introducing imperfections (Prescott (1986)).

At the time, macroeconomics looked (and felt) more divided than ever before (the intensity of the debate is well reflected in Lucas and Sargent (1978)). Yet, nearly 20 years later, the two routes have surprisingly converged. The methodological contributions of the Real Business Cycle approach, namely the development of stochastic dynamic general equilibrium models, have proven important and have been widely adopted. But the initial propositions that money did not matter, that technological shocks could explain fluctuations, and that imperfections were not needed to explain fluctuations, have not held up. The empirical

17 As it was absent from general equilibrium theory, leading economists working on stability and formalizations of real time tatonnement processes into similar dead ends.

evidence continues to strongly support the notion that monetary policy affects output. And the idea of large, high frequency movements in the aggregate production function remains an implausible black box; the relation between output and productivity appears more likely to reflect reverse causality, with movements in output leading to movements in measured total factor productivity, rather than the other way around.

For those reasons, most if not all current models, in either the New Keynesian or the New Classical mode (these two labels will soon join others in the trash bin of history of thought) now examine the implications of imperfections, be it in labor, goods, or credit markets. This is the body of work that I now turn to.

4 POST 1980. I. WORKING OUT THE QUANTITY THEORY.

In discussing the role of imperfections in macroeconomics, it is useful to divide the set of questions into two:

- The old Quantity Theory questions: Why does money affect output? What are the origin and the role of nominal rigidities in the process? These may be the central questions of macroeconomics, not because shifts in money are the major determinant of fluctuations (they are not), but because the non neutrality of money is so obviously at odds with the predictions of the benchmark, flexible price model.
- The old Business Cycle questions: What are the major shocks that affect output? What are their propagation mechanisms? What is the role of imperfections in that context?

This section focuses on the first set of questions, the next section on the second set.

Most macroeconomists would, I believe, agree today on the following description of what happens in response to an exogenous increase in nominal money:¹⁸

- Given the price level, an increase in nominal money leads to an increase in the aggregate demand for goods. At given nominal prices – and so at given relative prices – this translates to an increase in the demand for each good. ('Prices' is used generically here. I do not distinguish between wages and prices).
- Increasing marginal cost/disutility implies that this increase in the demand for each good leads each price setter to want to increase his price relative to others.
- If all individual prices were set continuously, the attempt by each price setter to increase his price relative to others would clearly fail. All prices, and by

¹⁸ Most but not all. While other explanations, based for example on distribution ('limited participation') effects of open market operations, have until now turned out to be dead ends, a number of macroeconomists remain skeptical of nominal rigidities as the source of the real effects of money.

implication, the price level would rise until the price level had adjusted in proportion to the increase in nominal money, demand and output were back to their original level, and there was no longer any pressure to increase relative prices. This would happen instantaneously. Money would be neutral, even in the short run.

- Individual prices however are adjusted discretely rather than continuously, and not all prices are adjusted at the same time. This discrete, staggered adjustment of individual prices leads to a slow adjustment of the average level of prices – the price level.¹⁹
- During the process of adjustment of the price level, the real money stock remains higher, and aggregate demand and output remain higher than their original value. Eventually, the price level adjusts in proportion to the increase in nominal money. Demand, output, and relative prices return to their original value. Money is neutral, but only in the long run.

This story feels simple and natural. Indeed, it is not very different from the account given by the quantity theorists of the 19th century. What the recent research has done has been to clarify various parts of the argument, and point to a number of unresolved issues.

4.1 *Staggering, and the Adjustment of the Price Level*

A tempting analogy to the proposition that staggered adjustment of individual prices leads to a slow price level adjustment is to the movements of a chain gang. Unless gang members can coordinate their movements very precisely, the chain gang will run slowly at best. The shorter the length of the chain between two gang members, or the larger the number of members in the gang, the slower it is likely to run.

Research on the aggregate implications of specific price rules and staggering structures has shown that the analogy is typically right. In most cases, discrete adjustment of individual prices leads indeed to a slow adjustment of the price level.²⁰ And the more each desired price depends on other prices, the slower the adjustment. But this research has also come with a number of warnings. In a celebrated counterexample to the general proposition, Caplin and Spulber (1987) have shown that, under some conditions, the reverse proposition may in fact hold. Discrete adjustment of individual prices may still lead to a completely flexible

¹⁹ An earlier hypothesis for slow adjustment of individual prices, developed by Lucas (1973) and based on incomplete information rather than staggering, has been largely abandoned. It is perceived to rely on implausible assumptions about the structure of information on macroeconomic aggregates.

²⁰ The most influential model here is surely Taylor (1980). See Taylor (1998) for a recent survey.

price level.²¹ The conditions under which their conclusion holds are more likely to be satisfied at high inflation, and this has an important implication. The effects of money on output are likely to be shorter, the higher the average rate of money growth and the associated rate of inflation.

4.2 *Real and Nominal Rigidities*

If individual price changes are staggered, the price level will increase only if at least some individual price setters want to increase their relative price. Once the price level has fully adjusted to the increase in money, and demand and output are back to their original level, desired relative prices end up the same as they were before the increase in money; but this is true only in the end.

This observation has one important implication. The speed of adjustment of the price level depends on the elasticity of desired relative prices in response to shifts in demand. The higher this elasticity, the more each individual price setter will want to increase his price when he adjusts, the faster the price level will increase, the shorter will be the effects of money on output. Research suggests that, to generate the slow adjustment of the price level one observes in the data, this elasticity must indeed be small, smaller than one would expect if for example the price set by firms reflected the increase in marginal cost for firms, and the wage reflected the increase in the marginal disutility of work for workers.²²

This proposition is sometimes stated as follows: ‘Real rigidities’ (a small elasticity of the desired relative price to shifts in demand) are needed to generate substantial ‘nominal rigidity’ (a slow adjustment of the price level in response to changes in money). The terminology may be infelicitous. But the conclusion is an important one and points to an interaction between nominal rigidities and other imperfections. If these other imperfections are such as to generate real rigidities (a big if), they can help explain the degree of nominal rigidity we appear to observe in modern economies.

4.3 *Demand versus Output*

Suppose that the price level responds slowly, so an increase in money leads to an increase in the demand for goods for some time. In the absence of further information on the structure in goods and labor markets, there is no warranty that this increase in demand will lead to an increase in output. It will do so only if suppliers of both labor and goods are willing to supply more. This was indeed the main unresolved issue in the fixed price equilibrium approach. For increases in

21 Their result requires two conditions: That prices be changed according to an Ss rule, and that each desired nominal price be a non decreasing function of time. Caplin and Leahy (1991) show what happens when only the first condition holds. Money is then typically non neutral.

22 See Blanchard and Fischer (1989), chapter 8, and Chari et al. (1998) for a recent discussion.

demand to translate into increases in output, the market structure must be such that the price setters are willing to supply more even at the existing price.

There are market structures where this will be the case. Suppose for example that the goods market is composed of monopolistically competitive price setters. At the initial equilibrium, monopoly power implies that their price is above their marginal cost. This implies in turn that, even at an unchanged price, they will be willing to satisfy an increase in demand, at least so long as the marginal cost remains smaller than the price. So, under this market structure, shifts in demand, so long as they are not too large, will lead to an increase in output.

For this reason, a typical assumption in recent research has been one of monopolistic competition. In the goods market, the assumption that price setters have some monopoly power and so may be willing to increase output in response to a shift in demand seems indeed reasonable. The assumption of monopolistic competition in the labor market is clearly much less appealing, and the question of whether the same conclusion will derive from more realistic descriptions of wage setting remains open. More generally, this points to yet another interaction between nominal rigidities and other imperfections. To understand why output and employment respond to shifts in demand, we need a better understanding of the structure of both goods and labor markets.

4.4 Money as Numeraire and Medium of Exchange

The argument underlying non-neutrality relies on two properties of money, money as the medium of exchange and money as the numeraire.

Staggering of individual price changes delivers slow adjustment of the average price of goods in terms of the numeraire. If, in addition, the numeraire is also the medium of exchange – which it need not be as a matter of logic, but typically is – slow adjustment of the average price of goods in terms of the numeraire means slow adjustment of the average price of goods in terms of the medium of exchange. It is these two features which, when combined, imply that changes in the stock of nominal money lead to changes in the value of the stock of money in terms of goods, leading in turn to movements in the interest rate, the demand for goods, and output.

This coincidence is crucial to the story. It suggests that delinking the numeraire and the medium of exchange would lead to very different effects of changes in nominal money – and more generally, of shifts in the demand for goods – on output. Put more strongly, it might change the nature of short run fluctuations. The idea of having a numeraire separate from the medium of exchange is an old idea in macroeconomics, an idea explored by Irving Fisher in particular. But, despite some recent work (for example Shiller (1999)), we still lack a good understanding of what macroeconomic implications, and by implication what potential benefits, could come from such a separation.

The set of results I have described above is sometimes called the ‘menu costs’ explanation of the short-run non-neutrality of money.²³ The expression correctly captures the notion that small individual costs of changing prices can have large macroeconomic effects. At the same time, the expression may have been a public relations disaster. It makes the explanation for the effects of money on output look accidental, when in fact the effects appear to be intrinsic to the workings of an economy with decentralized price and wage setting. In any economy with decentralized price and wage setting, adjustment of the general level of prices in terms of the numeraire is likely to be slow relative to a (fictional) economy with an auctioneer.

5 POST 1980. II. THE ROLE OF OTHER IMPERFECTIONS

Leaving aside nominal rigidities, there are three main reasons why macroeconomists working on fluctuations should care about imperfections:²⁴

- Imperfections lead to very different efficiency and welfare characteristics of the equilibrium, and thus modify the way we think about fluctuations and the role of policy. Think for example of the question of whether the equilibrium rate of unemployment is too high or too low, and its implications for macroeconomic policy.²⁵
- Imperfections may lead to very different propagation mechanisms of shocks. Think for example of the role of the interactions of nominal and real rigidities in determining the persistence of changes in money on output I discussed in the previous section.
- Imperfections may lead to new sources of shocks. Think for example of bank runs, which may affect not only the supply of money, but also the functioning of the financial intermediation system, leading to long-lasting effects on output.

These are the motivations behind the research on imperfections and macroeconomics which has developed in the last 20 years or so. As I indicated in the introduction, this phase is still very much one of exploration. The thousand flowers are still blooming, and it is not clear what integrated macroeconomic model will emerge, if any. Let me describe four major lines along which substantial progress has already been made.

23 See in particular Mankiw (1985) and Akerlof and Yellen (1985).

24 The arguments would be even stronger if the focus of this article were extended to cover growth. Much of the recent progress in growth theory has come from looking at the role of imperfections and institutions (externalities from R&D and patent laws, bankruptcies and bankruptcy laws, restrictions to entry by new firms, institutions governing corporate governance, etc.) in growth.

25 For example, the implications for monetary policy emphasized by Barro and Gordon (1983).

5.1 *Unemployment and the Labor Market*

The notion that the labor market was somehow special was reflected in early Keynesian models by the crude assumptions that the nominal wage was given, and employment was determined by the demand for labor. It was reflected in the confused debates about whether unemployment was involuntary or voluntary. It was reflected by the continuing use of an *ad hoc* formalization of wage behavior – the Phillips curve – even in theoretical models. It was reflected by the unease with which the neo-classical formulation of labor supply, developed by Lucas and Rapping (1969), with its focus on intertemporal substitution, was received by most macroeconomists.

For a long time, the basic obstacle was simply how to think of a market where, even in equilibrium, there were some unsatisfied sellers – there was unemployment. The basic answer was given in the early 1970s, in a set of contributions to a volume edited by Phelps (1970): One should think of the labor market as a decentralized market, in which there were workers looking for jobs, and firms looking for workers. In such a market, there would always be, even in equilibrium, both some unemployment and some vacancies.

Research started in earnest in the 1980s, based on a number of theoretical contributions to search and bargaining in decentralized markets, in particular by Diamond (1982b), Mortensen (1982), and Pissarides (1985). The conceptual structure that has emerged is known as the flow approach to the labor market:²⁶

- The labor market is a decentralized market. At any point in time, because of shifts in the relative demand for goods, or changes in technology, or because a firm and a worker no longer get along, a number of employment relations are terminated, and a number of new employment relations are started.

The workers who separate from firms, because they quit or are laid off, look for new jobs. The firms, which need to fill new jobs or replace the workers who have left, look for workers. At any point in time, there are both workers looking for jobs – unemployment – and firms looking for workers – vacancies.

- When a worker and a firm meet and conclude that the match seems right, there is typically room for bargaining.

The wage is then likely to depend on labor market conditions. If unemployment is high, and vacancies are low, the firm knows it will be able to find another worker easily, and the worker knows it will be hard to find another job. This is likely to translate into a low wage. If unemployment is low, vacancies high, the wage will be high.

- The level of the wage in turn affects the evolution of both vacancies and unemployment. A high wage means more terminations, less starts, and so a decrease in vacancies and an increase in unemployment. Conversely, a low wage

²⁶ For recent surveys, see Pissarides (2000), or Mortensen and Pissarides (1999).

leads to an increase in vacancies and a decrease in unemployment. Given these dynamics, the economy typically converges to a set of equilibrium values for unemployment and vacancies. One can then think of the natural rate of unemployment as the value to which the unemployment rate converges.

It is clear that in such an economy, there should and always will be some unemployment (and some vacancies). Operating the economy at very low unemployment would be very inefficient. But the equilibrium level of unemployment typically has no claim to being the efficient level. The equilibrium level, and its relation to the efficient level, depend on the nature of the process of search and matching, as well as on the nature of bargaining between workers and firms.

One way of assessing progress is to go back to a famous quote by Friedman (1968) about the natural rate of unemployment:

‘The natural rate of unemployment is the level which would be ground out by the Walrasian system of general equilibrium equations, provided that there is imbedded in them the actual structural characteristics of the labor and commodity markets, including market imperfections, stochastic variability in demands and supplies, the cost of gathering information about job vacancies and labor availabilities, the costs of mobility, and so on.’

What we have done is to go from this quote to a formal framework in which the role of each of these factors (and many others) can be understood and then taken to the data.²⁷ Today, we have a much better sense of the role and the determinants of job creation and job destruction, of quits and layoffs, of the nature of the matching process between firms and workers, of the effects of labor market institutions from unemployment benefits to employment protection, on unemployment. This line of research has shown for example how higher employment protection leads to lower flows in the labor market, but also to longer unemployment duration. Lower flows and longer duration unambiguously change the nature of unemployment. But because, in steady state, unemployment is the product of flows times duration, together they have an ambiguous effect on the unemployment rate itself. Both the unambiguous effects on flows and duration are indeed clearly visible when looking across countries with different degrees of employment protection.

Many extensions of this framework have been explored. While the basic approach focuses on the implications of the specificity of the product being transacted (labor services by a particular worker to a particular firm), and the room for bargaining that this creates, a number of contributions have focused on other dimensions, such as the complexity of the product being transacted, and the information problems this raises. For example, how much effort a worker puts in his job can be hard to monitor for a firm. If the firm just paid this worker his

27 For a more detailed assessment, see Blanchard and Katz (1997).

reservation wage, he would not care about being found shirking and being fired. One way the firm can induce him not to shirk is by paying him a wage higher than his reservation wage, so as to increase the opportunity cost of being found shirking and being fired. This particular effect has been the focus of an influential paper by Shapiro and Stiglitz (1984), who have shown that, even absent issues of matching, the implication would be positive equilibrium unemployment. As they have argued, in this case, equilibrium unemployment plays the role of a (macroeconomic) 'discipline device' to induce effort on the part of workers.

So far however, our improved understanding of the determinants of the natural rate of unemployment has not translated into a much better understanding of the dynamic relation between wages and labor market conditions. In other words, we have not made much progress since the Phillips curve. In particular, current theoretical models appear to imply a stronger and faster adjustment of wages to labor market conditions than is the case in the data. One reason may be the insufficient attention paid to yet another dimension of labor transactions, namely that they typically are not spot transactions, but rather long-term relations between workers and firms. Long-term relations often allow for better outcomes, for reasons emphasized by the theory of repeated games. For example, they may allow the wage to play less of an allocational role, more of a distributional or insurance role.²⁸ Here may lie one of the keys to explaining observed wage behavior. That firms might want to develop a reputation for good behavior and, by so doing, achieve a more efficient outcome, is indeed one of the themes of the research on 'efficiency wages.' But after much work in the 1980s, this direction of research has tapered off, without the emergence of a clear picture or an integrated view.²⁹ This is a direction in which more work is clearly needed.

5.2 *Saving, Investment, and Credit*

Issues of financial intermediation, from 'credit crunches' to 'liquidity scrambles,' figured heavily in the early accounts of fluctuations.³⁰ With the introduction of the *IS-LM* model, the focus shifted away. Issues of financial intermediation were altogether absent from the *IS-LM* model and most of its descendants. The treatment of financial intermediation in larger models was often schizophrenic. Asset markets were typically formalized as competitive markets, with a set of arbitrage relations determining the term structure of interest rates and stock prices. Among financial intermediaries, typically only banks, because of their relevance to the determination of the money supply, were treated explicitly. Credit problems were dealt with implicitly, by allowing for the presence of current cash flow in invest-

28 See for example Espinosa and Rhee (1989).

29 For a survey of work up to 1986, see Katz (1986).

30 See, for example, the 1949 survey by McKean.

ment decisions, and of current income in consumption decisions.³¹ One can therefore see recent research as returning to old themes, but with better tools (asymmetric information) and greater clarity.

The starting point, when thinking about credit, must be the physical separation between borrowers and lenders. Lending means giving funds today in anticipation of receiving funds in the future. Between now and then, many things can go wrong. The funds may not be invested but squandered. They may be invested in the wrong project. They may be invested, but not paid back. This leads potential lenders to put limits on how much they are willing to lend, and to ask borrowers to put some of their own funds at risk. How much borrowers can borrow therefore depends on how much cash or marketable assets they have to start with, and on how much collateral they can put in – including the collateral associated with the new project.

This fact alone has a number of implications for macroeconomic fluctuations:

- The distribution of income and marketable wealth between borrowers and lenders matters very much for investment.
- The expected future matters less, the past and the present – through their effect on the accumulation of marketable assets by firms – matter more for investment. Transitory shocks, to the extent that they affect the amount of marketable assets, can have lasting effects. Higher profits today lead to a higher cash flow today, and thus more investment today and more output in the future, a mechanism emphasized by Bernanke and Gertler.
- Asset values play a different role than they do in the absence of credit problems. Shocks which affect the value of marketable assets can affect investment even when they do not have a direct effect on future profitability – a channel explored for example by Kiyotaki and Moore (1997).

In such a world, the importance of having marketable assets leads in turn to a demand for marketable, or ‘liquid,’ assets by consumers and firms. Because consumers find it hard to either insure against idiosyncratic income shocks, or to borrow against future labor income, consumers save as a precaution against adverse shocks (Carroll (1997), Deaton (1992)). Here, theoretical and empirical research on saving have made clear that both life cycle and precautionary motives play an important role in explaining saving behavior. Similar considerations apply to firms. Because firms may need funds to start a new project, or to continue with an existing project, they also have a demand for marketable assets, a demand for liquidity. A new set of general equilibrium questions arises: Will the

31 A notable exception here is the work by Eckstein and Sinai (summarized in Eckstein and Sinai (1986)), and reflected in the DRI model. With its focus on balance sheets of firms and intermediaries, it was surprisingly at odds with the language and the other models of the time. But many of its themes have since come back into fashion.

economy provide such marketable assets? Can the government for example improve things by issuing such liquid instruments as T-bills?³²

In such a world also, there is clearly scope for financial intermediaries to alleviate information and monitoring problems. But their presence raises a new set of issues. To have the right incentives, financial intermediaries may need to have some of their own funds at stake. Thus, not only the marketable assets of ultimate borrowers, but also the marketable assets of intermediaries matter. Shocks in one part of the economy which decrease the value of their marketable assets can force intermediaries to cut on lending to the rest of the economy, resulting in a credit crunch (Holmström and Tirole (1997)). And to the extent that financial intermediaries pool funds from many lenders, coordination problems may arise. An important contribution along these lines is the clarification by Diamond and Dybvig (1983) of the nature and the necessary conditions for bank runs.

Because some of their liabilities are money, banks play a special role among financial intermediaries. In that light, recent research has looked at and clarified an old question, namely whether monetary policy works only through interest rates (the standard ‘money channel’), or also by affecting the amount of bank loans and cutting credit to some borrowers who do not have access to other sources of funds (the ‘bank lending’ channel).³³ The tentative answer at this point: The credit channel probably played a central role earlier in time, especially during the Great Depression, but, because of changes in the financial system, its importance may now be fading.

Research on credit is one of the most active lines of research today in macroeconomics. Much progress has already been made. This is already reflected, for example, in the (*ex post*) analyses of the Asian crisis, and in the analysis of the problems of financial intermediation in Eastern Europe.

Let me turn to the two remaining themes. I shall be more succinct because less progress has been made, in the first case because the evidence remains elusive, in the second because much remains to be done.

5.3 *Increasing Returns*

The potential relevance of increasing returns to fluctuations is another old theme in macroeconomics. The argument is straightforward:

- If increasing returns to scale are sufficient to offset the short-run fixity of some factors of production such as capital, short-run marginal cost may be constant or even decreasing with output. Firms may be willing to supply more goods at roughly the same price, leading to longer lasting and larger effects of shifts in

32 See for example Holmström and Tirole (1998).

33 Bernanke and Gertler (1995) give a general discussion of the way credit market imperfections can modify the effects of monetary policy on output.

aggregate demand on output (a version of the interaction between nominal and real rigidities discussed earlier).

- Indeed, if returns are increasing enough, the economy may have multiple equilibria, a low-efficiency low-output equilibrium and a high-efficiency high-output equilibrium. Many examples of such multiple equilibria have been worked out. Some have been based on increasing returns in production (Kiyotaki (1988)), others on increasing returns in exchange (Diamond (1982a)).

A related line of research has focused on countercyclical mark-ups (of price over marginal cost). Just like increasing returns, countercyclical mark-ups may explain why firms are willing to supply more output at roughly the same price. Like increasing returns, countercyclical mark-ups imply that the economy may operate more efficiently at higher output, in this case not because productivity is higher but because distortions (the gap between price and marginal cost) are lower. A number of models of mark-ups, based on imperfect competition, have been developed, some indeed predicting countercyclical mark-ups (for example, Rotemberg and Woodford (1991)), others however predicting procyclical mark-ups (Phelps (1992)).

Turning to the empirical research gives the same feeling of ambiguity. It appears to be true that, in response to exogenous shifts in demand, firms are willing to supply more at nearly the same price (for example Shea (1993)). How much is due to flat marginal cost or to countercyclical mark-ups is still unclear. Countercyclical mark-ups appear indeed to play a role (for example Bilal (1987)), but the source of such countercyclicality (nominal rigidities, lags in the adjustment of prices to cost, changes in the desired or in the sustainable mark-up if the mark-up is thought to result from a game among oligopolistic firms) remains to be established. For the time being, the possibility of multiple equilibria, due either to increasing returns or countercyclical mark-ups, remains an intriguing but unproven hypothesis.

5.4 *Expectations as Driving Forces*

This last theme, movements in expectations as an important source of fluctuations, is once again an old one. It was a dominant theme of pre-1940 macroeconomics. It was a major theme in Keynes, and beyond. But, with the introduction of rational expectations in the 1970s, expectations became fully endogenous, and the theme was lost.

To most macroeconomists, rational expectations is the right benchmark. But this only means that the burden of the proof is on those who insist on the presence of deviations from rational expectations, on their relevance for asset prices, and in turn for output fluctuations. This is indeed where a lot of research on 'behavioral finance' has recently taken place.

Research has taken place along two fronts.³⁴ The first has looked at the way people form expectations. A substantial body of empirical – often experimental – evidence has documented that most people form their expectations in ways which are not consistent with the economists’ definition of full rationality, for example in ways inconsistent with Bayes’ rule. Some sequences of realizations are more ‘salient’ than others, and have more effect on expectations than they should. For example, there is some evidence that a sequence of positive past returns leads people to revise expectations of future returns more than they should. This appears potentially relevant in thinking about phenomena such as fads or bubbles in asset markets.

For deviations from rationality by some investors to lead to deviations of asset values from fundamentals, another condition must be satisfied: There must be only limited arbitrage by the other investors. This is the second front on which research has advanced. The arguments it has made are simple ones. First, the required arbitrage is typically not riskless: What position do you take if you believe the stock market is overvalued by $x\%$? Second, professional arbitrageurs do not have unlimited funds. This opens the same issues as those we discussed earlier when discussing credit. Asymmetric information implies a limit on how much these arbitrageurs can borrow, and thus on how much they can arbitrage incorrect asset prices.³⁵

Empirical research has shown that this line of research can account for a number of apparent puzzles in asset markets. But other explanations, not based on such imperfections, may also account for these facts. At this stage, the question is far from settled. At issue is a central question of macroeconomics, the role of expectations, of bubbles and fads, as driving forces for at least some macroeconomic fluctuations.

6 MACRO IN THE (NEAR) FUTURE

Let me briefly restate the thread of my argument. Relative to Wicksell and Fisher, macroeconomics today is solidly grounded in a general equilibrium structure. Modern models characterize the economy as being in temporary equilibrium, given the implications of the past, and the anticipations of the future. They provide an interpretation of fluctuations as the result of shocks working their way through propagation mechanisms. Much of the current work is focused on the role of imperfections.

34 For a review, see Shleifer (2000).

35 One of the small victories of this line of research has been the description of an LTCM-type crisis before it happened. In 1997, Shleifer and Vishny argued that it was precisely at the time when asset prices deviated most from fundamentals that arbitrageurs might be least able to get the external funds they needed to arbitrage and may need to liquidate their positions, making things worse. This is what happened one year later at LTCM.

What happens next? Predicting the evolution of research is very much like predicting the stock market. Like financial arbitrage, intellectual arbitrage is not perfect, but it is close. Let me nevertheless raise a number of questions and make a few guesses.

Which imperfections? Part of the reason current research often feels confusing comes from the diversity of imperfections invoked in explaining this or that market. To caricature, but only slightly: Research on labor markets focuses on decentralization and bargaining; research on credit markets focuses on asymmetric information; research on goods markets on increasing returns; research on financial markets on psychology.

To some extent, the differences in focus must be right. The problems involved in spot transactions are different from those involved in intertemporal ones; the problems involved in one-time transactions are different from those involved in repeated transactions, and so on. Still, if for no other than aesthetic reasons, one may hope for an integrated macro model, based on only a few central imperfections (say, those which give rise to nominal rigidities, and one or two others).

There exists indeed a few attempts to provide such an integrated view. One is by Phelps in 'Structural Slumps' (1994), which is based on implications of asymmetric information in goods and labor markets. Another is by Caballero and Hammour (for example (1996)), based on the idea that most relations, either in credit or in labor markets, require some relation-specific investment, and therefore open room for hold ups *ex post*. These are important contributions, but I see them more as the prototype cars presented in car shows but never mass-produced later. They show what can be done, but they are probably not exactly what will be.

Policy and welfare. One striking implication of recent models is how much more complex the welfare implications of policy are. To take one example, in a model with monopolistic competition, (small) increases in output due to the interaction of money and nominal rigidities improve welfare to a first order. This is because initially marginal cost is below the price and the increase in output reduces the wedge between the two. Under other distortions, the effects of output fluctuations on efficiency and welfare can be much more complex. Recent research has revisited for example the question of whether recessions 'cleanse' the economy – by eliminating firms which should have been closed anyway – or weaken it – by destroying perfectly good firms. The answer so far is: Both, in proportions which depend on the precise nature of imperfections in labor and credit markets. This clearly has implications for how one views fluctuations.³⁶

The medium run. There has been a traditional conceptual division in macroeconomics between the short run – the study of business cycles – and the long run

36 See for example Caballero and Hammour (1999).

– the study of growth.³⁷ A better division might actually be between the short run, the medium run, and the long run. Phenomena such as the long period of high unemployment in Europe in the last 25 years, or the behavior of output during the transition in Eastern Europe, do not fit easily into either business cycles or growth. They appear to involve different shocks from those generating business cycles – changes in the pace or the nature of technological progress, demographic evolutions, or in the case of Eastern Europe, dramatic changes in institutions. They also appear to involve different imperfections – with imperfections in labor, credit and goods markets rather than nominal rigidities playing the dominant role – than business cycles.

Research on imperfections has allowed us to make substantial progress here. Our understanding of the evolution of European unemployment for example is still far from good; but it is much better than it was ten or twenty years ago.

Macroeconomics and institutions. The presence of imperfections typically leads to the emergence of institutions designed, more or less successfully, to correct them. Examples range from anti-trust legislation, to rules protecting minority shareholders, to unemployment insurance. Which institutions emerge is clearly important in understanding medium run evolutions. Think of the role of labor market institutions in explaining European unemployment, the role of legal structures in explaining the evolution of output in transition economies. Institutions also matter for short run fluctuations, with different institutions leading to different shocks and propagation mechanisms across countries. For example, a recent paper by Johnson et al. (1999) argues that, during the recent Asian crisis, those Asian countries that had the weakest governance institutions (such as poor protection of minority shareholders) were also those that suffered the largest exchange rate declines. Identifying the role of differences in institutions in generating differences in macroeconomic short and medium-run evolutions is likely to be an important topic of research in the future.

Current debates. In the early 1980s, macroeconomic research seemed divided in two camps, with sharp ideological and methodological differences. Real business cycle theorists argued that fluctuations could be explained by a fully competitive model, with technological shocks. New Keynesians argued that imperfections were of the essence. Real business cycle theorists used fully specified general equilibrium models based on equilibrium and optimization under uncertainty. New Keynesians used small models, capturing what they saw as the essence of their arguments, without the paraphernalia of fully specified models.

Today, the ideological divide is gone. Not in the sense that underlying ideological differences are gone, but in the sense that trying to organize recent con-

³⁷ More than the rest of this essay, this reflects my views rather than some assessment of the consensus. See for example Blanchard (1997).

tributions along ideological lines would not work well. As I argued earlier, most macroeconomic research today focuses on the macroeconomic implications of some imperfection or another. At the frontier of macroeconomic research, the field is surprisingly a-ideological.

The methodological divide is narrower than it was, but is still present. Can macroeconomists use small models such as the *IS-LM* model or the Taylor model – the model of aggregate demand and supply with wage staggering developed by John Taylor? Or should they use only fully specified dynamic general equilibrium models, now that such models can be solved numerically?³⁸ Put in these terms, it is obvious that this is an incorrectly framed debate. Intuition is often obtained by playing with small models. Large explicit models allow for further checking and refining. Small models then allow again to convey the essence of the argument to others. At this stage, I believe that small models are indeed underused and undertaught.³⁹ Small, back-of-the envelope, models are much too useful to disappear, and I expect that methodological divide will also fade away.

One way to end is to ask: Of how much use was macroeconomic research in understanding, for example, and helping resolve the Asian crisis of the late 1990s?

Macroeconomists did not predict either the time, place, or scope of the crisis. Previous exchange rate crises had involved either fiscal misbehavior (as in Latin America), or steady real appreciation and large current account deficits (as in Mexico). Neither fiscal policy, nor, given the very high rate of investment, the current account position of Asian countries seemed particularly worrisome at the time.⁴⁰

So, when the crisis started, macroeconomic mistakes (such as fiscal tightening, the right recommendation in previous crises, but not in this one) were made. But, fairly quickly, the nature of the crisis was better understood, and the mistakes corrected. And most of the tools needed were there to analyze events and help the design of policy, from micro-based models of bank runs, to models of financial intermediation, to variations on the Mundell-Flemming model allowing, for example, for an effect of foreign currency denominated debt on the balance sheet, and in turn on the behavior of firms and banks.

Since then, a large amount of further research has taken place, leading to a better understanding of the role of financial intermediation in exchange rate crises. Based on this research, proposals for better prudential regulation of financial institutions, for restrictions on some forms of capital flows, and for a redefinition

38 This debate is about models used in research, not about the applied econometric models used for forecasting or policy.

39 Paul Krugman recently wondered how many macroeconomists still believe in the *IS-LM* model. The answer is probably that most do, but many of them do not know it well enough to tell.

40 A notable exception here is the work of Calvo (for example Calvo (1998)), who, before the crisis took place, emphasized the potential for maturity mismatch (short term foreign debt, long term domestic loans) to generate an exchange rate crisis, even absent fiscal or current account deficits.

of the role of the IMF are being discussed. A passing grade for macroeconomics? Given the complexity of the issues, I think so. But it is for the reader to judge.

REFERENCES

- Akerlof, G. and J. Yellen (1985), 'A Near-rational Model of the Business Cycle with Wage and Price inertia,' *Quarterly Journal of Economics*, 100 (Supplement), pp. 823–838.
- Barro, R. and H. Grossman (1976), *Money, Employment, and Inflation*, Cambridge University Press, Cambridge.
- Barro, R. J. and D. Gordon (1983), 'A Positive Theory of Monetary Policy in a Natural Rate Model,' *Journal of Political Economy*, 91(4), pp. 589–610.
- Bernanke, B. and M. Gertler (1989), 'Agency Costs, Net Worth and Business Fluctuations,' *American Economic Review*, 79, pp. 14–31.
- Bernanke, B. and M. Gertler (1995), 'Inside the Black Box: The Credit Channel of Monetary Policy Transmission,' *Journal of Economic Perspectives*, 9(4), pp. 27–48.
- Bils, M. (1987), 'The Cyclical Behavior of Marginal Cost and Price,' *American Economic Review*, 77, pp. 838–855.
- Blanchard, O. (1997), 'The Medium Run,' *Brookings Papers on Economic Activity*, 2, pp. 89–158.
- Blanchard, O. and S. Fischer (1989), *Lectures on Macroeconomics*, MIT Press.
- Blanchard, O. and L. Katz (1997), 'What Do We Know and Do Not Know about the Natural Rate of Unemployment,' *Journal of Economic Perspectives*, 11(1).
- Caballero, R. and M. Hammour (1996), 'The 'fundamental transformation' in macroeconomics,' *American Economic Review*, 86(2), pp. 181–186.
- Caballero, R. and M. Hammour (1999), 'The Cost of Recessions Revisited: A Reverse-Liquidationist View,' Mimeo, MIT.
- Calvo, G. (1998), 'Varieties of Capital Market Crises,' in: G. Calvo and M. King (eds.), *The Debt Burden and its Consequences for Monetary Policy*, McMillan.
- Caplin, A. and J. Leahy (1991), 'State-dependent Pricing and the Dynamics of Money and Output,' *Quarterly Journal of Economics*, 106, pp. 683–708.
- Caplin, A. and D. Spulber (1987), 'Menu Costs and the Neutrality of Money,' *Quarterly Journal of Economics*, 102, pp. 703–726.
- Carroll, C. (1997), 'Buffer-stock Saving and the Life Cycle/Permanent Income Hypothesis,' *Quarterly Journal of Economics*, 112(1), pp. 1–56.
- Chari, V., P. Kehoe, and E. Mc Grattan (1998), 'Sticky Price Models of the Business Cycle: Can the Contract Multiplier Solve the Persistence Problem?,' Staff Paper 217, Federal Reserve Bank of Minneapolis.
- De Wolff, P. (1941), 'Income Elasticity of Demand, a Micro-Economic and a Macro-Economic Interpretation,' *Economic Journal*, LI, pp. 140–145.
- Deaton, A. (1992), *Understanding Consumption*, Oxford University Press.
- Diamond, D. and P. Dybvig (1983), 'Bank Runs, Deposit Insurance, and Liquidity,' *Journal of Political Economy*, 91, pp. 401–419.
- Diamond, P. (1982a), 'Aggregate Demand Management in Search Equilibrium,' *Journal of Political Economy*, 90, pp. 881–894.
- Diamond, P. (1982b), 'Wage Determination and Efficiency in Search Equilibrium,' *Review of Economic Studies*, 49, pp. 217–227.

- Dornbusch, R. (1976), 'Expectations and Exchange Rate Dynamics,' *Journal of Political Economy*, 84, pp. 1161–1176.
- Eckstein, O. and A. Sinai (1986), 'The Mechanisms of the Business Cycle in the Postwar Era,' in: R. Gordon (ed.), *The American Business Cycle. Continuity and Change*, NBER and the University of Chicago Press, pp. 39–122.
- Espinosa, M. C. and Rhee (1989), 'Efficient Wage Bargaining as a Repeated Game,' *Quarterly Journal of Economics*, 104-3, pp. 565–588.
- Fisher, I. (1911), *The Purchasing Power of Money*, MacMillan, New York.
- Friedman, M. (1968), 'The Role of Monetary Policy,' *American Economic Review*, 58, pp. 1–21.
- Frisch, R. (1965), 'Propagation Problems and Impulse Problems in Dynamic Economics,' in: *Readings in Business Cycles*, Irwin, New York, pp. 155–185. First published in 1933.
- Haberler, G. (1937), *Prosperity and Depression. A Theoretical Analysis of Cyclical Movements*, League of Nations, Geneva.
- Hall, R. (1978), 'Stochastic Implications of the Life-cycle Permanent Income Hypothesis: Theory and Evidence,' *Journal of Political Economy*, 86(5), pp. 971–987.
- Hansen, A. (1951), *Business Cycles and National Income*, Norton, New York.
- Hicks, J. (1937), 'Mr Keynes and the Classics,' *Econometrica* 5.
- Hicks, J. (1939), *Value and Capital*, Oxford University Press.
- Holmström, B. and J. Tirole (1997), 'Financial Intermediation, Loanable Funds, and the Real Sector,' *Quarterly Journal of Economics*, 112, pp. 663–692.
- Holmström, B. and J. Tirole (1998), 'Private and Public Supply of Liquidity,' *Journal of Political Economy*, 106(1), pp. 1–40.
- Johnson, S., P. Boone, A. Breach, and E. Friedman (1999), 'Corporate Governance in the Asian Financial Crisis, 1997-1998,' Mimeo, MIT, January.
- Kahn, R. (1931), 'The Relation of Home Investment to Unemployment,' *Economic Journal*, 41, pp. 173–198.
- Katz, L. (1986), 'Efficiency Wage Theories: A Partial Evaluation,' *NBER Macroeconomics Annual*, 1, pp. 235–276.
- Keynes, J. (1964), *The General Theory of Employment, Interest, and Money*, Harcourt, New York, first published 1936.
- Kiyotaki, N. (1988), 'Multiple Expectational Equilibria under Monopolistic Competition,' *Quarterly Journal of Economics*, 102(4), pp. 695–714.
- Kiyotaki, N. J. and Moore (1997), 'Credit Cycles,' *Journal of Political Economy*, 105(2), pp. 211–248.
- Klein, L. (1946), 'Macroeconomics and the Theory of Rational Behavior,' *Econometrica*, 14(2), pp. 93–108.
- Lucas, R. (1973), 'Some International Evidence on the Output-Inflation Trade-off,' *American Economic Review*, 63(3), pp. 326–334.
- Lucas, R. (1987), *Models of Business Cycles*, Basil Blackwell, Oxford.
- Lucas, R. and L. Rapping (1969), 'Real Wages, Employment, and Inflation,' *Journal of Political Economy*, 77, pp. 721–754.
- Lucas, R. and T. Sargent (1978), *After Keynesian Economics, After the Phillips Curve: Persistence of High Inflation and High Unemployment*, Federal Reserve Bank of Boston.
- Lucas, R. and N. Stokey (1989), *Recursive Methods in Economic Dynamics*, Harvard University Press, Cambridge.

- Mankiw, N. (1985), 'Small Menu Costs and Large Business Cycles: A Macroeconomic Model,' *Quarterly Journal of Economics*, 100(2), pp. 529–539.
- Marshall, A. (1879), *The Economics of Industry*, MacMillan, London.
- Marshall, A. (1923), *Money, Credit and Commerce*, MacMillan, London.
- McKean, R. (1949), 'Liquidity and a National Balance Sheet,' *Journal of Political Economy*, pp. 506–522.
- Metzler, L. (1951), 'Wealth, Saving, and the Rate of Interest,' *Journal of Political Economy*, 59, pp. 93–116.
- Mitchell, W. (1923), *Business Cycles and Unemployment*, National Bureau of Economic Research, McGraw-Hill, New York.
- Modigliani, F. (1944), 'Liquidity Preference and the Theory of Interest and Money,' *Econometrica*, 12(1), pp. 45–88.
- Modigliani, F. and R. Brumberg (1954), 'Utility Analysis and the Consumption Function: An Interpretation of Cross-section Data,' in: K. Kurihara (ed.), *Post Keynesian Economics*, Rutgers University Press, pp. 388–436.
- Mortensen, D. (1982), 'The Matching Process as a Noncooperative/Bargaining Game,' in: J.J. McCall (ed.), *The Economics of Information and Uncertainty*, Chicago, University of Chicago Press, pp. 233–254.
- Mortensen, D. and C. Pissarides (1999), 'Job Reallocation, Employment Fluctuations, and Unemployment Differences,' in: John Taylor and Michael Woodford (eds.), *Handbook of Macroeconomics*, chapter 18, Elsevier Science, pp. 1171–1182.
- Obstfeld, M. and K. Rogoff (1996), *Foundations of International Macroeconomics*, MIT Press.
- Patinkin, D. (1965) (first edition, 1956), *Money, Interest, and Prices. An Integration of Monetary and Value Theory*, Harper and Row, New York.
- Phelps, E. (1970), *Microeconomic Foundations of Employment and Inflation Theory*, W.W. Norton, New York.
- Phelps, E. (1992), 'Customer Demand and Equilibrium Unemployment in a Working Model of the Incentive-Wage Customer-Market Economy,' *Quarterly Journal of Economics*, pp. 1003–1033.
- Pigou, A. (1936), 'Review of *The General Theory of Employment, Interest, and Money* by J.M. Keynes,' *Economica*, 3, pp. 115–132.
- Pigou, A. (1950), *Keynes' General Theory: A Retrospective View*, MacMillan, London.
- Pissarides, C. (1985), 'Short-run Equilibrium Dynamics of Unemployment, Vacancies, and Real Wages,' *American Economic Review*, 1985, pp. 676–690.
- Pissarides, C. (2000), *Equilibrium Unemployment Theory*, second edition, MIT Press, forthcoming.
- Prescott, E. (1986), 'Theory ahead of Business Cycle Measurement,' *Quarterly Review, Minneapolis Fed*, pp. 9–22.
- Ramsey, F. (1928), 'A Mathematical Theory of Saving,' *Economic Journal*, 38(152), pp. 543–559.
- Rotemberg, J. and M. Woodford (1991), 'Markups and the Business Cycle,' in: Olivier Blanchard and Stanley Fischer (eds.), *NBER Macroeconomics Annual*, pp. 63–128.
- Samuelson, P. (1939), 'Interactions between the Multiplier Analysis and the Principle of Acceleration,' *Review of Economic Statistics*, 21(2), pp. 75–78.
- Sargent, T. (1973), 'Rational Expectations, the Real Rate of Interest, and the Natural Rate of Unemployment,' *Brookings Papers on Economic Activity*, 2, pp. 429–472.
- Sargent, T. (1987), *Dynamic Macroeconomic Theory*, Harvard University Press, Cambridge.
- Shapiro, C. and J. Stiglitz (1984), 'Equilibrium Unemployment as a Discipline Device,' *American Economic Review*, 74, pp. 433–444.

- Shea, J. (1993), 'Do Supply Curves Slope Up?,' *Quarterly Journal of Economics*, 108(1), pp. 1–32.
- Shiller, R. (1999), *Designing Indexed Units of Account?*, NBER Working Paper 7160.
- Shleifer, A. (2000), *Inefficient Markets. An Introduction to Behavioral Finance*, Clarendon Lecture, Oxford University Press, Oxford U.K.
- Shleifer, A. R. and Vishny, R. (1997), 'The Limits of Arbitrage,' *Journal of Finance*, 53, pp. 35–55.
- Snowdon, B. H. and Vane (1999), *Conversations with Leading Economists. Interpreting Modern Macroeconomics*, Edward Elgar, Cheltenham, U.K.
- Solow, R. (1999), 'Interview,' in: B. Snowdon and H. Vane (eds.), *Conversations with Leading Economists. Interpreting Modern Macroeconomics*, Edward Elgar, Cheltenham, U.K.
- Taylor, J. (1980), 'Aggregate Dynamics and Staggered Contracts,' *Journal of Political Economy*, 88(1), pp. 1–24.
- Taylor, J. (1998), 'Staggered Price and Wage Setting in Macroeconomics,' NBER WP 6754, forthcoming *Handbook of Macroeconomics*.
- Tinbergen, J. (1939), *Statistical Testing of Business Cycle Theories*, League of Nations, Geneva, volume II.
- Wicksell, K. (1898), *Interest and Prices*, MacMillan, London, first published in German in 1898.
- Wolff, P. De (1941), 'Income Elasticity of Demand, a Microeconomic and a Macroeconomic Interpretation,' *Economic Journal*, 51(201), pp. 140–145.
- Woodford, M. (1999), 'Revolution and Evolution in Twentieth-century Macroeconomics,' mimeo.

