1. Introduction

Capital is at the heart of capitalism: it is, accordingly, not surprising that we should look to failures in capital markets to account for one of the most important failures of capitalism, the marked fluctuations in output and employment which have characterized capitalism throughout its history.

For more than a decade now, I and several of my co-authors (in particular, Bruce Greenwald and Andrew Weiss) have been exploring the thesis that it is imperfections in the capital market – imperfections which themselves can be explained by imperfect information – which account for many of the peculiar aspects of the behavior of the economy which macroeconomics attempts to explain.¹ In this lecture, I want to provide an overview of that endeavor, to explain why it is that we have increasing confidence in the importance of capital market imperfections, to contrast our

¹Others before us have recognized the importance of capital markets. Perhaps the most complete articulation of these views which is closest in spirit to that presented here is that of Lindbeck (1963), who goes on to stress the distinction between debt and equity (which, as we shall see, underlies much of our analysis), and the role of credit rationing. More recent studies include the work of Gerler and Bernanke (1989, 1990). Our own work is presented in a series of papers (see references). A set of recent empirical studies are collected together in Hubbard (1990).

Capital markets played an important role in the ideas of Leijonhufvud and Minsky, and in the later work of Hicks (1988). The role of capital market imperfections in investment was, ironically, stressed in one of the earliest econometric studies of investment, that of Mayer and Kuh (1959).
2. The anomalies

Science has sometimes been described as an attempt to explain anomalies, things which are puzzling. Economic science is no different. Among the principal puzzles which have preoccupied macro-economists, five will be the center of my attention today.

2.1. The magnitude of the fluctuations

The first puzzle concerns the magnitude and persistence of the fluctuations which have buffeted the economy (table 1). It is not just that one sector has its ups and downs, but that the aggregate level of activity, whether measured by employment or output, fluctuates, and its principal components, including aggregate consumption and investment, fluctuate, and fluctuate together. To be sure, the world economy is constantly buffeted by shocks – good weather in Iowa, leading to a bumper crop of maize there; bad weather in Brazil,
leading to a small coffee harvest. But macro-economics is concerned with aggregate movements, and the puzzle is, why should these sectoral shocks make much difference: to a large extent they should be offsetting — shifts in demand result in positive boosts to one sector, offset by declines in others. Moreover, standard price theory predicts that price responses will dampen the magnitudes of the overall economy’s responses: an initial supply shock, for instance, results in a much smaller quantity response as a result of an offsetting price movement.

Lags and buffers serve to dampen further the economy’s response: the fluctuations in coffee output may have hardly any consequences on the permanent income of coffee growers, or the real income of coffee drinkers; a bad crop necessitates using up some of the coffee put away in a buffer stock; an exceptionally good crop entails putting coffee into the buffer stock. For households, savings provide a buffer; individuals can draw out of the buffer in bad years, and put into the buffer in good years. Even in the rapidly changing field of computers, it will be several years before the new invention which has made my computer obsolete is fully disseminated — long enough that the economy may have turned half or more of a cycle.

It is precisely the recognition of the difficulty of finding real shocks that could account for the magnitude of observed fluctuations that led so many economists to look to monetary disturbances.

2.2. The non-neutrality of money

But here, we encounter the second puzzle: the classical dichotomy said that money should not matter; money determined the price level, but real magnitudes — output, employment, investment, etc. — were determined by relative prices (including relative intertemporal prices, i.e. real interest rates). Yet the evidence suggests clearly that monetary policy has real effects. Monetary policy might affect money prices, and even money interest rates (intertemporal money prices), but why should it affect anything real? Keynesian and New Keynesian economists have a ready answer: price rigidities. But, as I shall explain below, the traditional explanations for those price rigidities remain shaky, the mechanisms by which price rigidities,

\[^2\]Indeed, recent theoretical work seems to have strengthened the case for the classical dichotomy; see, e.g. Barro (1974) and Stiglitz (1988b).

\[^3\]Too abundant to cite here: but I have in mind more than just the money-income correlations, which, as Tobin forcefully pointed out, are hardly convincing. Most neutral observers should have been convinced of the non-neutrality of monetary policy by the major recession which the Fed induced in the United States in the early 1980s.

\[^4\]See below for some of the critiques of the currently most fashionable explanation, ‘menu costs’.
combined with changes in the money supply, lead to real effects remain
debatable,5 and, overall, the conclusion that these price rigidities, combined
with fluctuations in the money supply, are to blame for the economic
fluctuations is even more suspect. It is not that one cannot write down
models in which such rigidities can have real effects – one obviously can –
but that one cannot formulate models in which the magnitudes of the effects,
through the specified channels, seem large enough to account for the non-
neutralities which are observed.

One reason that this hypothesis is suspect is the worldwide experience of
output and employment fluctuations in the presence of inflation: surely, the
problems in this case cannot be attributed to downward money rigidities.
Another reason is that if we look more closely at some of the loci of
fluctuations – investment, and in particular, construction (see table 1) – we
see that these sectors, in which prices are particularly flexible (prices are
negotiated, since items are produced individually) are even more volatile than
aggregate output.

2.3. Shifts in the aggregate supply curve

So far, we have argued that the observed fluctuations in output are greater
than can be accounted for by standard theory: given that many of the shocks
to the economy are offsetting, given the (price) shock absorbers, buffers, and
lags built in the economy, it is hard to find real disturbances of the required
magnitude; and it is hard to understand how monetary disturbances have any
effect, let alone effects of the required magnitude. Whatever the source of the

5One long popular theory (which underlies many of the fixed price models) employs real
balance effects. Changes in real balances of households makes them feel wealthier, and thus
consume more; more telling than the theoretical criticisms, based on the Barro-Ricardo effect, is
the fact that, given empirical estimates of the wealth elasticity of consumption, this effect simply
is not very important, even with the largest observed variations in real balances.

A second mechanism focuses on the money market; an increase in the real supply of money
lowers interest rates; elsewhere [Stiglitz (1988a)], I have provided an extensive critique of this
mechanism; the most important point is simply that, over long periods of time, variations in real
interest rates have simply been too small for them to play an important role.

Finally, Tobin has emphasized the effect of changes in the real money and bond supply on the
price of equities, through general portfolio effects. At a theoretical level, these models (in the
absence of capital market imperfections, the subject of this lecture) do not provide a convincing
theory [see Stiglitz (1988b)]: empirically, models testing Tobin's q have not fared well, and for
good reason – as we point out below, new equity is a relatively unimportant source of funds for
investment. The observed correlation is more than likely not a causal relation: stock prices and
investment are both high when firm prospects are good.

Though one can easily construct models where changes in the rules by which the money
supply changes have real effects (even when the mean rate of change of the money supply is
constant) when households are risk averse, again these effects almost surely do not account for
the magnitude of the observed non-neutralities.
disturbance, there is a third puzzle: to a large extent, the output variations represent shifts in the aggregate supply curve. What can account for such large shifts in the aggregate supply schedule?

Consider a simple aggregate production function, where, as we do conventionally, we keep the magnitude of capital fixed in the background:

\[ Q = F(L), \]

(1)

where \( Q \) is aggregate output and \( L \) is aggregate employment. Then labor ought to be hired up to the point where the real wage equals the marginal product,

\[ \frac{w}{p} = F'(L), \]

(2)

where \( w \) is the wage and \( p \) the price level; or, viewing this from the supply side, output is a function simply of the real wage.

\[ Q = F(F^{-1}(\frac{w}{p})). \]

(3)

Real (product) wages do not vary much over the business cycle, so that output produced should not vary much. In fact, as table 2 illustrates, in many sectors, real product wages actually fell in the Great Depression, implying that output should have increased! The rigidity of real wages, while itself a puzzle, makes the fluctuations in output all the more puzzling. What can account for the evidently large shift in the aggregate supply schedule?

### 2.3.1. Changes in the degree of competition

There are two resolutions of this puzzle which I wish to dispense with.
quickly. The first is that there is imperfect competition, and that changes in the degree of monopoly can account for the observed shifts in output. Of course, with imperfect competition, we need to dispense with the aggregate supply curve. We replace the condition (2) with the condition

$$w/p = (1 - 1/\beta)F'(L),$$

where $\beta$ is the elasticity of demand, so that aggregate output, as a function of the real wage, can be written

$$Q = F(F'^{-1}(w/p(1 - 1/\beta))).$$

If competition becomes less keen in a recession, then at a given real wage, output will be reduced. While I believe that product markets are imperfectly competitive [within the macro-economics context, Robert Hall (1988) has provided convincing evidence on this], and that these imperfections do play a role (though a secondary one) in any complete explanation of the economy's fluctuations, I do not believe that the variations in the degree of competition can account for the shifts in the (pseudo-) supply functions. This is not to deny that it is conceivable – even likely – that the degree of competition varies over the cycle. The variety of reasons for this I explored in my 1984 paper. The question is, can the pattern and magnitudes of these variations account for the observed pattern and magnitudes of variations of real product wages and output in different industries? One problem is that we should expect the degree of competition to decrease in some industries, as, for instance, the threat of entry becomes decreased; but to increase in others, as the value of collusion, particularly of firms whose life expectancy has been reduced, decreases. More importantly, among the industries in which we see the most variability are those, like construction, that appear most competitive. Surely, in these industries, oligopoly models cannot explain pricing policies.

2.3.2. Technology shocks (real business cycles)

The second explanation for the shifting supply curve is that technology is changing. Obviously, if we rewrite the production function

$^6$Okun's Law can be interpreted as having long made the case for some kind of market imperfection: the fact that output increases more than proportionately with employment means that the economy exhibits short run increasing returns. This is consistent, for instance, with monopolistic competition, with firms operating on the downward sloping portion of their average cost curve. The claim that Okun's Law can be explained by some kind of labor hoarding does not obviate the necessity of invoking some form of imperfection in competition: for labor hoarding implies that the short run marginal cost of production is very low.

$^7$Though admittedly, if collusion is to be sustained, it may necessitate an increase in mark-ups.

$^8$For instance, of the kind explored by Rotemberg and Saloner (1986).
where $A$ represents the 'state of technology', then (5) becomes

$$Q = AF(F^{-1}(w/pA)),$$

so variations in $A$ will indeed induce variations in $Q$ at fixed real wages. But we cannot account for the magnitude of those fluctuations, in particular for the downward movements in the supply function. While crop failures may be interpreted as a negative shock to $A$, the downward movement of the economy in the Great Depression, or more recently in the Reagan recession or the Bush recession, cannot be traced to any such shock. Yet, remarkably, a major school of macro-economics within the United States – the real business cycle theories, which I discuss further below – has attempted to do just that.

2.4. **Patterns of sectoral movements**

The fourth puzzle is concerned with patterns of sectoral movement, some aspects of which I have already referred to. Investment in general, and inventories and construction in particular, fluctuate more than output (table 1). The puzzle of these fluctuations is that in recessions, the marginal costs of construction or production are low. Real interest rates are sufficiently low that we ought to observe production smoothing – inventory accumulation in recessions, decumulation in booms. We ought to see firms take advantage of the low marginal costs to build up inventories, even in recessions.

9The list of criticisms of the technological interpretation to economic fluctuations is not meant to be exhaustive: if the source of disturbance were technological, then the shocks should be industry related, and correlations in movements of the outputs of particular industries across countries should be greater than correlations of movements of output across industries within a country. This appears not to be the case.

Nor can the theory really do an adequate job accounting for the magnitude of the fluctuations: given the buffers, lags, and the price 'shock absorber' the original disturbances would have to be truly large, and their occurrence would be associated with large changes in capital values. It is hard to find the direct empirical confirmation of these large disturbances. Surely, if they were large, they should be observable!

10See Blinder and Macchi (1991). As they point out, there have been a number of attempts to resolve the inventory paradox within the neoclassical framework, but none has really succeeded. Some inventories (such as inventories of raw materials) are used for production purposes, and these should accordingly increase in proportion to output. The inventory puzzle with which we are concerned is most apparent in the consumer goods sector. Lucas (1981) discussed the broader puzzle of volatility in investment: 'The volatility of business investment over the cycle is at least as severe a paradox as the cyclical behavior of employment, since the principal characteristic of optimal investment behavior (as we understand it theoretically) is the way it smooths reactions to transient shocks'. He goes on to provide an explanation: 'Volatile cyclical investment must be explicable, exactly as it [sic] volatile cyclical employment, only as a repeated mistake' (p. 15). The model we present shows that one does not have to rely on mistaken expectations.
of low construction costs to build long-lived capital goods, the value of which is likely to be virtually unaffected whether the recovery begins this month or in six months from now.

2.5. Persistence

The fifth, and final, puzzle to which I wish to draw attention is that of persistence, the serial correlation in output and employment. It is easy to construct models that give rise to some persistence, even if the economy were buffeted with i.i.d. shocks. Any model in which there is a state variable (like capital) will give rise to some correlation: the effects of a negative shock will be felt in subsequent years. The problem is not to construct a model with some correlation, but to construct plausible models with correlations of the required magnitudes.

3. Some preliminary methodological remarks

There is a basic, well recognized scientific principle we have employed in our exploration of explanations for these puzzles: Ockham’s razor. We have sought a single explanation (or set of related explanations) which explain all of these phenomena, or at least as many as possible. We have based our analysis on micro-economic principles, not the micro-economic principles of neoclassical models based on a complete set of perfectly competitive markets with perfect information; but rather the micro-economic principles derived from models in which markets are incomplete and information is imperfect (with the incompleteness of markets being largely accounted for by informational imperfections).

The pseudo-scientific ‘as if’ methodology promoted by Friedman has been pushed too far: a theory is to be tested not by a selected set of its predictions, but by all of its predictions. Among the testable parts of a theory are those assumptions which are themselves directly testable. Assumptions of technological regression are testable, and, I dare say, provide by themselves grounds for rejecting the real business cycle theory. We count it a virtue, not a vice, if the assumptions of the model are themselves plausible; if the micro-foundations underlying the theory are themselves testable – and tested.

Our approach should be contrasted not only with real business cycle theory but also with the piecemeal approach taken in some recent macro-economic work. One theory (the menu cost theory) explains price rigidity by referring to the cost of adjustment of prices; another theory explains labor hoarding by referring to the cost of adjustment of labor. Having explained wage and price rigidities, still another theory is needed to explain the shifts in the supply curve; and still another theory to explain lay-offs. The piecemeal approach is not only unattractive, in requiring a myriad of explanations, but the explanations seem sometimes at odds with one another.
Fig. 1. Menu costs and other costs of adjustment. If the demand curve shifts, the firm must either adjust prices and/or quantities. The menu cost literature seems to assume that there are large costs to adjusting $p$, but no costs to adjusting quantities.

If there are costs of adjusting outputs and inputs and prices, clearly all should be taken into account by the firm in deciding on how to react to a shift in its demand curve. Fig. 1 shows a shift in the demand curve. The menu cost theory, focusing on the cost of adjusting prices, says that the firm will keep prices constant, adjusting output, moving from $Q_0$ to $Q_1$. But this makes sense only if there is no cost of adjusting output. If there is a cost of adjusting output – as surely there is – and if these costs exceed the costs of adjusting prices (the small menu costs), as seems plausibly the case in most industries – then the response will be a relatively fixed output and large variations in prices, such as standard neoclassical theory would predict.

4. Our central hypotheses

Our analysis begins from the premise that capital markets – the markets through which capital and risks are transferred from one party to another – are imperfect. Though we, and others, have related these capital market imperfections to costly and imperfect information and transactions costs, it

Table 3
Role of equities as a source of finance.

<table>
<thead>
<tr>
<th>Period, country</th>
<th>Net equity issues as a percent of total investment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-financial enterprises</td>
<td></td>
</tr>
<tr>
<td>Canada (1970–1985)</td>
<td>2.5</td>
</tr>
<tr>
<td>Finland (1970–1985)</td>
<td>-0.1</td>
</tr>
<tr>
<td>France (1970–1985)</td>
<td>-0.4</td>
</tr>
<tr>
<td>Germany (1970–1989)</td>
<td>0.9</td>
</tr>
<tr>
<td>Italy (1970–1985)</td>
<td>8.2</td>
</tr>
<tr>
<td>Japan (1970–1987)</td>
<td>2.7</td>
</tr>
<tr>
<td>1946–1958, U.S.</td>
<td>6.4</td>
</tr>
<tr>
<td>1901–1912, U.S.</td>
<td>14</td>
</tr>
<tr>
<td>1923–1929, U.S.</td>
<td>19</td>
</tr>
<tr>
<td>1949–1953, U.S.</td>
<td>6</td>
</tr>
<tr>
<td>1949–1982, Electrical engineering firms, U.K. (shares plus long term liabilities)</td>
<td>15.0</td>
</tr>
</tbody>
</table>


is our belief that the source of the imperfection is not so important as its nature and consequence. In particular, our analysis is based on two hypotheses.

(1) The distinction between equity and debt. While debt imposes a particularly heavy risk burden on the firm (since the firm must absorb all fluctuations in profits), imperfections in the equity market are severe, so that a relatively small fraction of new capital is raised by new equity issues. (See table 3 and fig. 2.)

(2) The distinctive nature of the mechanisms by which credit (and capital more generally) is allocated. Credit is not allocated by means of an auction/price process; in many cases, it is rationed; information is spread disparately throughout the economy; there is limited transferability of information; and financial institutions, including banks, play a central role in the allocation process.
Our analysis proceeds by several steps. Focusing initially on the first hypothesis, we analyze the behavior of the firm. The firm has a portfolio of interrelated decisions to make. Among the portfolio of decisions facing a firm are its price and wage decisions, and its employment and investment decisions. We describe how these decisions respond to changes in attributes of the firm (in particular its net worth) and the economic environment in which it operates (in particular, its perception of the risks which it faces). These individual firm decisions are then aggregated to describe the aggregate behavior of the economy.

In this first stage, we assume that the firm is not credit constrained. The firm has limited access to the equity market, but not to the credit market. Our results are thus a consequence of the form in which capital is available, not the amount – a distinction which our first hypothesis stresses. In the second stage of our analysis, we explore the consequences of limited credit availability, and the determinants of its availability.

5. Equity rationing and risk averse firms

The first hypothesis, combined with the assumption of significant bank-
ruptcy costs, has one critical implication: firms act in a risk averse manner. They cannot divest themselves of risk through equity markets. Higher levels of investment or production entail increased debt; the fixed obligations entailed by debt imply that as debt is increased, the bankruptcy probability is increased.\footnote{In some formulations, if firms have a sufficiently low net worth, they act in a risk loving manner. The circumstances in which this occurs and the consequences are discussed more fully below.}

5.1. Modelling risk averse behavior

We have modelled the risk averse firm behavior in two different ways: maximizing expected terminal wealth, minus expected bankruptcy costs,\footnote{See, for instance, Greenwald and Stiglitz (1988c, 1992).}

\[
\max E(a(x, z), B) - cP_B,
\]

where \( a \) is terminal value of the firm, a function of \( \pi \), profits, which in turn are a function of \( x \), a set of decision variables, and \( z \), a set of (for the moment) exogenous parameters (environmental variables), and of \( B \), the amount the firm has borrowed (which in turn is a function of \( x \), the decision variables, and \( a_0 \), the initial net worth of the firm), \( c \) is the cost of bankruptcy, and \( P_B \) is the probability of bankruptcy, which in turn is a function of \( x \), the decision variables, \( \bar{r} \), the interest rate the firm must pay, and the firm's initial net worth. The interest rate the firm must pay, in turn, is a function of the decision variables \( x \), the initial wealth \( a_0 \), and the 'safe' rate of interest, \( \rho \). In Greenwald and Stiglitz (1992) we show how, in the case of risk neutral lenders with rational expectations, we can solve simultaneously for \( \bar{r} \) and \( P_B \):

\[
\bar{r} = r(x; a_0; z; \rho), \quad P_B = P(x; a_0; z; \rho).
\]

Alternatively,\footnote{See, for instance, Greenwald and Stiglitz (1987a, 1989a).} we can think of firms as maximizing

\[
EU(a),
\]

where \( U \) is a concave function of profits.

5.1.1. Dynamic formulation

Of course, we need to view the firm within a dynamic framework, with decisions today affecting the state variables which it passes on to the future.
The three most important state variables are 'cash', $a$, inventories, $N$, and capital stock, $K$:15

$$a_{t+1} = p_{t+1} Q[p_{t+1}, \mu_1] - B_t(1 + \hat{r}) - d_t,$$  \hspace{1cm} (10)

where

$$B_t = w_t L_t + I_t - a_t,$$  \hspace{1cm} (11)

$$K_{t+1} = K_t + I_t - \mu_2 K_t,$$  \hspace{1cm} (12)

and

$$N_{t+1} = N_t + \mu_3 \Phi(L_t, K_t) - Q[p_{t+1}, \mu_1]$$  \hspace{1cm} (13)

and where $B =$ debt, $\Phi$ is the production function, $Q$ now represents sales, $d_t =$ dividends,16 and $\mu_t$ is a vector of random variables ($\mu_1$ is a random variable affecting demand, $\mu_2$ the depreciation rate, and $\mu_3$ production). Within this more general framework, we replace the objective function (8) or (9) with a dynamic programming problem

$$\max u(d_t) + J(a_{t+1}, K_{t+1}, N_{t+1}),$$

where $J(a, K, N)$ is the terminal valuation of the asset position. Much of the discussion will, for simplicity, be couched in terms of the simpler objective function (8), or, alternatively, (9); reports using (14) are similar.

### 5.2. Basic results

The first central result of our analysis is that firms produce up to the point where (in the first formulation) expected marginal returns equal expected marginal bankruptcy costs,17 which, with risk aversion, will be at a lower level than that at which expected (net) returns are zero.

Consider the special case of a single decision variable, the level of employment. Firms, in our theory (the first formulation), produce up to the point where

$$pF' - w = \psi,$$

where $\psi$ is the marginal bankruptcy cost (fig. 3). If we ignored the latter, we

15In fact, any variable which cannot costlessly and instantaneously be adjusted should be treated as a state variable. From this perspective, wages, prices, and employment should also be treated as state variables. On the other hand, if there were perfect rental markets and no costs of adjustment of capital, $K$ would not be a state variable, from the perspective of the firm, though it might be from the perspective of the economy as a whole.

16We cannot present here a theory of dividend behavior. Empirical studies indicate dividends are closely linked to firms' net worth, though they ratchet up, and exhibit downward rigidities. Such rigidities can be explained in terms of the signalling role that dividends (or, more generally, firm distributions) play. See Bhattacharya (1979). Our analytic work has employed the simpler hypothesis that $d_t$ is just a function of $a_t$.

17In the second formulation the expected marginal utility of (net) returns is zero.
Fig. 3. Output determination and volatility with risk. In the absence of risk, output (employment) is determined at the point where \( pF' - w = 0 \) (the value of the marginal product equals the wage). Since real wages do not change much in the short run, and since the production function does not shift much in the short run, output does not shift. With bankruptcy risk, employment is set where \( pF' - w = \psi \), where \( \psi \) is the marginal bankruptcy cost. This can shift markedly in the short run. It increases as a firm's net worth falls, and as the perceived riskiness of production increases.

would obtain the standard condition of firms producing up to the point where the value of the marginal product equals the wage. With little variability in the real wage, it is impossible to account for variability in supply (at the firm or macro-level); our third basic puzzle. We attribute shifts in the aggregate supply curve to shifts in \( \psi \), the (marginal) risk of bankruptcy. Actual bankruptcy rates vary markedly over time (see fig. 4), and it is plausible to assume that marginal bankruptcy probabilities do as well.

Having located the source of variability in the decisions undertaken by the firm, we need to look more closely at what determines shifts in \( \psi \).

5.2.1. Amplification and volatility

Changes in the net worth of the firm and the riskiness of the environment (the random variables) affect the decisions of the firm. This result may hardly seem striking, except when it is recalled that under the standard neoclassical
Looking at eq. (10), it is apparent that a small change in demand may result in a large change in \( a \): profits are a residual, and a 5% reduction in the price received or the quantity sold (below what was anticipated) may result in a 50% reduction in profits, if the profit margin is 10% (costs of production, including fixed interest payments, represent 90% of the price). Moreover, if firms are highly levered, a small change in \( a \) can result in large changes in the optimal value of decision variables, like output and employment. In one limiting case we explored, the change in output was proportionate to the change in \( a \): the result looked like a standard accelerator model, but applied not only to investment in fixed assets, but also to production. A 50% reduction in \( a \) would result in a 50% reduction in output. Though we think this extreme, the lesson is more general: small disturbances can lead to large effects. That is why our analysis has paid less attention to the source of the
disturbances to the economy than to their consequences: a downturn may be initiated by an oil price shock, a monetary shock, or a dramatic change in expectations. But the structure of the economy results in an amplification of the effects of these disturbances, in contrast to the prediction of standard theory, in which price shock absorbers, lags and buffers all serve to dampen the effects of disturbances.

The volatility of expectations, and the sensitivity of output to expectations (perceived risks) stressed by our theory provides a further channel by which the effects of shocks to the economy may be amplified.

The fact that risk and leverage differ across sectors provides part of the explanation for differences in volatility across sectors (other parts will be described below); the construction industry, for instance, is highly leveraged, and bankruptcy is not only common, it is particularly highly variable there.

5.2.2. Shifts in the aggregate supply function

It is apparent that when we aggregate, the aggregate supply function will be volatile, because of the volatility of changes in net worth and expectations. But there is a further reason for shifts in the aggregate supply function. Earlier, we noted that in the standard neoclassical model, many of the disturbances to the economy are offsetting; gains to one sector are offset by losses to another. This is far less true in our model.

Let us focus on the special case where the only asset of the firm is 'a' (ignoring inventories and capital). Under normal conditions, output is a concave function of a: \( Q_r = Q(a_r; \ ) \) (fig. 5). This implies that shocks to the economy that at first glance might appear to be nothing more than redistribution-shifts in demand that lead to higher profits by one firm and lower offsetting profits to another - have real effects. Starting from an initial equilibrium position, an oil price increase or an oil price decrease may have adverse effects on the economy. This implication is markedly different from the standard model; there, an oil shock increase has positive effects for oil producers, negative effects for oil users; and the net effect simply depends on which predominates. Also, in that theory if, in total, the net effect of a price increase is negative, then the net effect of a price decrease is positive.

5.2.3. Persistence

Furthermore, the dynamic processes described above entail a slow response to an initial disturbance: if, for some reason, net worth is reduced at time \( t \), production falls in subsequent periods, and only gradually will production be restored to normal. Fig. 6 shows the performance of the economy after a downward shock to \( a_r \). On average, taking into account the effect of the decision variables on the state variable, \( a_{t+1} \) the average value of \( a_{t+1} \) depends on the value of \( a_r \). The economy was initially at the long equilibrium, \( a^* \), where
Fig. 5. Output and firm net worth. Output is a concave, increasing function of net worth. Net worth matters, and concavity implies that changes in the distribution of net worth also matter. Increasing the dispersion of net worth (but keeping mean net worth unchanged) lowers average output from $Q^*$ to $Q$.

$$Ea_{t+1} = a_t = a^*.$$  

After the disturbance, with $a = a_0$, the economy only gradually gets restored to the long run equilibrium."}

5.2.4. Price and wage rigidities

Finally, the model implies different magnitudes of changes for different variables in response to changes in the environment, e.g. to changes in net worth. Abstracting, for a moment, from our more general dynamic model, consider, for instance, the simplest case where there are two decision variables, $x_1$ and $x_2$, with terminal wealth taking on the special form. 

---

18 This is not quite precise: the steady state is described by an ergodic state; a more precise statement is that only gradually does the probability distribution of observed states correspond to the ergodic state. We show that the difference equation for $a$ can give rise to chaotic cycles, in general equilibrium versions of the model where the demand for labor affects the wages, and hence profits. See Greenwald and Stiglitz (1992).

19 In this formulation, we assume the 'price' of $x_1$ is 1, and that the interest rate actually paid is random: it is $\hat{r}$ if the firm does not go bankrupt, and equals whatever the residual value of the firm dividend by the size of the outstanding debt if it does. $\hat{r}$ is set so that the expected value of $r$ is equal to the safe rate of interest.
Fig. 6. Persistence. After a negative shock to net worth (a movement from $a^*$ to $a_0$), the economy recovers only gradually.

$$a = f(x_1) + \mu g(x_2) - (1 + r)(a_0 - x_1 - x_2).$$

Under normal assumptions (decreasing absolute risk aversion) lower wealth involves a shift to safer activities: a shift out of inventories and into cash, and away from long run commitments, like investments and new employees.

The adjustment of some variables themselves are likely to entail more risk than others: it may be more difficult to predict the consequences of a change.

20 Using our formulation where firms maximize the expected utility of terminal wealth, the first order conditions are

$$EU'[f' - (1 + r)] = 0, \quad EU'[\mu g' - (1 + r)] = 0,$$

or

$$E[U'\mu]/EU' = f'/g'.$$

The reduction in $a$ can be viewed as a transformation of the utility function; with lower $a$, the individual acts in a more risk averse manner, provided only that the utility function is characterized by decreasing absolute risk aversion [see Rothschild and Stiglitz (1970, 1971) or Diamond and Stiglitz (1974)], i.e. $EU'\mu/EU'$ is smaller (with more weight placed on lower values of $\mu$), and hence the ratio of $f'$ to $g'$ is smaller, i.e. for any given value of $x_2$, $x_1$ is larger.
in wages and prices than of output. The effect of wage changes depends on how workers respond, and how employers at other firms respond. If they respond similarly, the adverse effects are likely to be much smaller than if they do not. The effect of price decreases depends on how customers respond, and how other producers respond. If, for instance, other firms respond similarly, the gains from price reductions are likely to be much smaller than if they do not. But predicting these responses is indeed difficult.

On the other hand, the worst that can happen if the firm cuts back its production is that its inventories will be slightly depleted, if sales turn out higher than expected; and the marginal cost of replenishing them may exceed the marginal cost of production today. The uncertainty associated with the effect on profits is only associated with the difference in marginal costs.

We argue that it is these differences in perceived risks which accounts for the relative inertia of wage and price decisions.21.22

5.2.5. The interdependence of supply and demand

Looking at the firms' decisions from a portfolio perspective has one further implication: we observe that investment and employment are, in a way, parallel decisions; they are both inputs into the production processes. One is an input focused on short-run output, the other affects output both in the short and the long run. New hires, requiring training, are even more closely analogous to an investment decision. But observing that investment and employment are closely related, we then see that, from the perspective of the macro-economy, demand and supply decisions are closely related: the variables that affect aggregate demand for investment are precisely the same variables that affect aggregate supply, the amount that firms are willing to

21 Indeed, not only may firms not lower their price, there is an argument that when the economy goes into a downturn, the firm may actually increase its price, at least relative to its marginal cost of production. In a world of imperfect competition, firms need to recruit customers. Recruiting customers is an investment (like the hiring decision). One way of affecting the rate at which customers are recruited is by lowering the price. Lowering the price loses profits today, but gains profits in the future. But as the economy goes into a downturn, the implicit cost of the lower price – the lower cash flow, inducing a higher probability of default – is increased, and the expected gain is reduced (since the firm's life expectancy is reduced). It pays firms to charge higher mark-ups.

22 Indeed, our analysis makes clear the fundamental incoherence of arguments that, in the face of certain adjustment costs, the firm should do nothing. With changing environments, they cannot do nothing. The argument of Akerlof and Yellen (1985) that, if firms have been maximizing with respect to wages or prices, the loss in profits from failing to adjust wages and prices is second order does not explain wage and price rigidities: for there is a second order loss from the failure to adjust any variable. Our analysis provides a natural definition of what it means to do 'nothing' – minimizing the risk of the firm.

The result that failure to adjust any variable, while it has a second order effect on the firm, has a first order effect on the economy (on social welfare) is a corollary of the Greenwald–Stiglitz theorem (1986), establishing that whenever the economy is not constrained Pareto efficient (which it will not be, in the stipulated circumstances) price adjustments (pecuniary externalities) can have first order effects.
produce. The above result concerning the relative magnitude of the variability of investment and employment may suggest that a disturbance to net worth may have a larger effect on aggregate demand than on aggregate short run supply, but the latter effect is significant, and cannot be ignored (fig. 7).

5.3. Explaining the basic puzzles

We have, with this first hypothesis, gone a long way to explaining the five anomalies which we set out to explain in the beginning. We have obtained large variations in output from relatively small disturbances. Small disturbances in demand can lead to large changes in a, and through that to large changes both in the demand for investment and the supply of goods. We have also explained the shift in the supply function, as well as the patterns of sectoral movements – the volatility of investment and inventories. And we have explained the fifth puzzle, the persistence of effects.23

23Our model can be used to explain other detailed characteristics of cyclical movements. For instance, as the economy moves out of a recession, firms first put workers on overtime before they hire new workers. If we used observed market rates of interest, this practice makes little sense, unless firms are extremely uncertain about whether the recovery is going to stick. Our model suggests that they not only need to be convinced about the recovery (so that the risk of investment in new workers is reduced), but also that their net worth has to be replenished (enhancing their capacity or willingness to undertake the risk of the investment). By the same
6. The availability of credit

So far, our analysis has focused on the supply of output and the demand for inputs (investment, employment). Associated with this, there is an implied demand for credit (the variable \( B \) in eq. 8). That analysis assumed an unlimited supply of credit at an (expected) return of \( \rho \). We now need to look more closely at the availability of credit. Our theory of credit availability is based on combining the principles derived earlier for the theory of risk averse firm to banks with the second hypothesis, which stresses the fact that credit is not allocated by means of an auction/price process. Our theory of credit availability will enable us to explain the one remaining puzzle, the money-output nexus.

6.1. Why price level matters with unindexed debt contracts

First, focusing for a moment on our hypothesis concerning the importance of the difference between debt and equity, we observe that if loan contracts are denominated in nominal interest rates, then any policy (including monetary policy) which affects the price level in an unexpected way has real effects on the economy. It generates real redistributions between debtors and creditors, between producers and households; in terms of our model, it leads to changes in the real value of \( a \), and we have seen how these changes in the average level as well as the distribution of \( a \) have real effects.\(^{24}\) To see how token, the reduction in wages which workers would have to take, to induce firms to bear the risks of investing in hiring them now, relative to what workers expect to obtain if they wait is such that workers are not willing to bear these costs.

Our model can also be used to explain the absence of profit sharing arrangements (workers become, effectively, equity participants, and the asymmetric information arguments for why equity markets do not work well also explain why workers may not value highly such equity participation.)

\(^{24}\) It is not an adequate attack against the theory to argue that we have not explained why loan contracts are not fully indexed. As the familiar joke puts it, the fact that some biologist has not explained (and indeed, might not be able to explain) why a giraffe has a long neck does not lessen the relevance of his taking into account the consequences of the length of neck. Surely, it is wrong to reason that because we cannot explain the length of the neck, or because according to some ill-conceived theory, rationally designed necks should be short, we should therefore proceed under the assumption that necks are indeed short. We should trust our eyes: giraffe necks are indeed long. And we should trust the evidence: most loan contracts are indeed not indexed. In fact it might be the case that if contracts were indexed some of the macroeconomic effects upon which we have focused would disappear. We can only conjecture about what a world might be like which is markedly different from the world in which we live.

It is, of course, a criticism of a theory to say that it is incomplete; and a theory which does not explain the lack of indexing is on that account indeed incomplete. Elsewhere, we and others have provided several suggestions concerning why contracts might not be indexed. There may be multiple equilibria: given that most contracts are not indexed, if a firm indexes one contract, he is left bearing more, not less, residual risk. Given asymmetries of information, innovations in contract form in bilateral negotiations are frequently looked upon askance [see Stiglitz (1990b)]: one party may be suspicious about the motives of the other for suggesting a reform in the standard contracts; with bounded rationality and limited information, the proposer may not be able to allay the suspicions that the reform is not meant to be a Pareto improvement, but an
changes in the price level may, in particular, affect the supply of credit, we need to look more closely at banks.

6.2. The theory of the risk averse bank

Banks are a specialized kind of firm, engaged in screening loan applicants, certifying creditworthiness and monitoring loans. While firms borrow money from banks, banks borrow money from their depositors. We can think of their production activity as making loans. This is a risky activity, and the general principles we derived earlier apply here: a reduction in the net worth of banks and an increase in the riskiness of their environment will lead them to contract their output, i.e. to make fewer loans.

We have established under a variety of conditions that when banks decide to lend less, they may not just put up the lower amount available to auction; rather they may ration credit. (For some purposes, the distinction is not that important: the consequences of markedly higher credit terms are much like those of credit rationing.)

The reduced availability of credit has multiplier effects throughout the economy. It has a direct effect on firms, who may not be able to invest as much as they otherwise would. And the reduced investment leaves its usual trail of effects.

But there are a set of indirect effects to which I want to call attention. We began this discussion by pointing out the parallel between banks and firms. Firms themselves are banks: they provide credit to their customers and suppliers. Given the localized nature of information (hypothesis 2), it makes sense for firms to ignore Polonius' injunction to Laertes, by 'both a borrower and a lender' being. Firms have considerable specialized knowledge about those with whom they trade. Lending too is a risky activity, which, when the access to credit of a firm is reduced (or the firm's net worth is reduced), is itself reduced: the firm reduces not only its investment in machines, but also in new customers, and in credit to its clients. Thus, the reduction in credit by a bank has a ripple effect on credit availability throughout the economy.

Even the anticipation of credit constraints has these effects: firms know that it takes time, and is costly, to adjust their asset positions; a threat of a reduction in credit availability in the future will lead them to attempt to enhance their cash position, cutting back on inventories and investment. Thus, rational expectations concerning future actions of the monetary authorities can have anticipatory effects; but unlike some models, where these attempt of one party to gain at the expense of the other. The legal system, with its presumption that if a non-standard contract is written, the terms of the contract be interpreted in favor of the party not writing the contract, reinforces the incentives against changes in standard contract forms.
rational expectations serve to offset the effects, here the anticipations may simply serve to increase the speed with which they take effect.

It should be clear now how, in the absence of the intervention of a monetary authority, an economic downturn can be amplified through the monetary-credit system. An economic downturn (regardless of its origins) results in higher than anticipated defaults, lowering banks' net worth. This, combined with the greater uncertainty associated with lending, reduces their lending activity, amplifying the economic downturn. Their reduced lending activity (possibly anticipated) leads firms to cut back on the credit they make available, on the level of investments in which they engage, the number of new employees they are willing to hire, and so the downturn gets amplified.

The death of a bank or a firm destroys localized knowledge about borrowers. It should not be surprising that there are only a limited number of firm/banks which have information about any borrower; for information is costly; it is a fixed, sunk cost – precisely characteristics which tend to preclude competition [see Jaffee and Stiglitz (1990)]. And the fact that information is not easily transferable means that the termination or reduction in lending activity from one source is not immediately offset by an increase in activity from other sources.

6.3. A simple model of bank behavior

A simple model captures some of the effects which we have just described. Fig. 8 shows a mean-variance curve for a risk averse bank. The locus \( R'R' \) shows the loan portfolio opportunities of the bank, given its net worth and deposits (which it is assumed to have no control over, with deposit interest rates being fixed). The loan locus is found by the bank maximizing over interest rates as well as other characteristics of lending policy. Alternatively,

\[ \text{This effect is offset, to some extent, in recessionary situations by price declines, which increase the real value of payments to the bank. But typically, the increased default rate of the recession more than offsets this effect.} \]

\[ \text{There is some ambiguity about what constitutes the 'death' of a firm or a bank. A reorganization entailing a change in ownership, but not in personnel, obviously does not destroy the on-going firm's information base. Lending activity is adversely affected by the reduction in the bank's net worth and by the reorganization process.} \]

\[ \text{The figure assumes there is credit rationing, and that the deposit rate is less than the T-bill rate. For other cases, see Greenwald and Stiglitz (1991b).} \]

\[ \text{The figure is constructed assuming all loan applicants are (in terms of their observable characteristics) identical. Credit rationing exists whenever the aggregate supply of loans (as derived in the figure) is less than the aggregate demand for loans. The existence of credit rationing does not depend on 'risk-loving' behavior of borrowers, induced by limited liability [in contrast to the impression sometimes conveyed to casual readers of Stiglitz-Weiss (1981)]. For instance, consider the case where a risk averse borrower is indifferent between two projects; the } \]

\[ \text{ith project yields a return of } R' \text{ with probability } p', \text{ and a zero return otherwise. Then an increase in the interest rate charged decreases expected utility by } U'[R'-(1+r)B]p'B, \text{ where } B \text{ is the amount borrowed; this is smaller for the project with the high return and low probability of success.} \]
the bank can invest in T-Bills. The over-all opportunity locus is the line $SR^*$. The bank chooses a point, such as $E$.

6.3.1. Net worth effects
A reduction in the bank's net worth shifts the locus $R'R'$ and the point $S$ in parallel. With decreasing absolute risk aversion, the amount of risk-taking – the size of the bank's loan portfolio – is reduced absolutely.

6.3.2. Financial fragility
As the economy goes into a recession, there are further effects: with an increased probability of default on the part of borrowers, risk is increased and expected return is reduced. The locus $R'R'$ shifts down and to the right. Substitution and wealth effects normally both lead to reduced risk taking; and even at the same degree of risk taking (as measured by the portfolio variance), there is less lending activity, since there is greater risk associated with any single loan. We call this the loan portfolio risk effect.

The shift in the loan portfolio locus $R'R'$ may be exacerbated by a worsening of the loan applicant pool. Earlier, we noted that with bankruptcy costs, firms act in a risk averse manner, provided that they have sufficient
capital. It has long been recognized that with limited liability, in the absence of bankruptcy costs, firms act as risk lovers. For any given bankruptcy cost, there is a critical net worth such that, below that net worth, firms act in a risk loving manner, above that, in a risk averse manner. Thus, as the economy moves into a recession, and firms find their net worth decreased, 'good' (i.e. risk averse) firms reduce their loan applications, 'bad' (i.e. risk loving) firms may increase their loan applications, and there is an increasing proportion of bad (i.e. low net worth) applicants. The lower net worth induces these firms to undertake even greater risks.

There are further reasons that the loan opportunity locus worsens. At all times, there is a turnover in a bank's customer portfolio, as some old customers go into bankruptcy, and new firms are created. In a recession, lending opportunities are shrinking, as regular customers go bankrupt, or move into the 'near-bankruptcy' category. A recession, with reduced bank net worth, is hardly the time to invest in new customer relations. There may be a presumption that a firm is seeking a new bank relationship has had its credit cut off by its existing bank (except in the case where its regular bank has gone into bankruptcy) – and the new bank knows that banks will be cutting off their worst prospects first. Firms are not in the process of borrowing to expand, but rather to survive.

We have thus identified several reasons that banks are likely to reduce lending activity as the economy goes into a recession:

(a) Reduced net worth leads to reduced risk taking, i.e. reduced lending.
(b) Worsening lending opportunities lead to reduced lending activity, as a result of wealth, substitution, and loan portfolio risk effects.
(c) The effect of bankruptcies of some banks is not offset by expansion of others.

And there are four reasons that lending opportunities appear to worsen as the economy goes into a recession:

(a) An increased probability of default by each loan applicant (the 'direct' increased risk effect);
(b) a decrease in the number of regular bank customers about whom the bank is informed (the decreased information effect);
(c) an increase in the proportion of risk loving applicants (an adverse selection effect);
(d) among this increasing proportion of risk loving applicants, an adverse

Firm's pay-offs are then max \( \{a, 0\} \), a convex function.
incentive effect, offset, to some extent, by a positive incentive effect among the shrinking proportion of risk averse applicants.

These effects may be so strong as to lead to a situation where the bank makes no loans at all: the credit market collapses, because the expected return to loans is actually lower than on T-bills. Since what is relevant for the curve $R'R'$ is lenders' perceptions of the riskiness of borrowers and their ability (based on their net worth) to bear risks, both of which can change quickly, it is evident that credit markets can exhibit considerable 'fragility'.

6.4. Monetary policy

How might monetary policy offset these effects? Can it? More generally, how does monetary policy work? In brief, monetary policy can affect the ability and willingness of banks to lend. Changing the reserve requirements obviously affects bank's ability to lend. In effect, in the United States, where reserves do not bear interest, increased reserve requirements move the opportunity locus $SR^*$ inward, in a parallel fashion: the wealth effect leads to less lending.

6.4.1. Open market operations

More subtle are the consequences of open market operations. These affect the supply of reserves, on the one hand, and the supply of T-bills being publicly traded on the other. We can cut into any general equilibrium system by looking at one market or another, and, in a complete description, it should make no difference. Here, we focus on the market for T-bills.

A reduction in supply of publicly available T-bills increases the price, and lowers interest rates. Depositor rates adjust slowly, more slowly than T-bill rates. Accordingly, changes in T-bill rates affect the 'seignorage' of banks. There is a wealth effect, which leads to less lending activity; but offsetting this is a substitution effect: low interest rates make T-bills less attractive investments for banks, relative to loans, and this by itself should encourage banks to lend more. (In terms of our diagram, fig. 9, an open market operation moves the point $S$ down relative to the curve $RR$.) When the bank is flush with capital, the threat of bankruptcy is low, wealth effects may be negligible, and the substitution effects may dominate: there is a strong

\[29\] Even when the expected return on loans exceeds that on T-bills, the bank may wish to make no loans. Our diagrams so far have assumed that there is no risk associated with T-bills. In fact, so long as T-bills are (imperfectly) indexed, there is real risk; and so long as there is real risk, the slope of the indifference curve at the point $S$ is strictly positive, and hence conceivably greater than the slope of the opportunity locus to the right of $S$. The opportunity locus will now not be a straight line; but the bank's decision, choosing a mix of an optimally chosen loan portfolio, and T-bills, is essentially the same [see the Cass–Stiglitz (1970) two-mutual fund theorem].
positive effect on loan supply. On the other hand, in bad times, when the bank is already worrying about bankruptcy, the wealth effect becomes stronger. In fig. 9, there is only a slight shift in the equilibrium to $E_1$.

Figs. 8–9 assumed that all loan projects (which passed the bank’s screening process) were identical. But some projects may be better than others, and fig. 10 shows the consequence, for an extreme case where there are only two types of projects. The opportunity locus then has a kink. The change in the T-bill rate shifts the opportunity locus, but the bank still chooses to operate at the kink. There are no additional loans made.

The moral of the story is simple: you can lead a horse to water but you can’t make it drink. In extreme recessions, it may be extremely difficult to induce banks to lend more.

Of course, the actual history of the economy is not traced out by the shifts in figs. 8–10. Rather, as we have noted, what is happening is that as the

---

$^{30}$There is also the possibility – made all too familiar by the S & L debacle in the U.S. – that if net worth becomes too low, the bank becomes a risk lover, and loan supply may actually increase. Effective bank regulation is designed to prevent this from happening.
Fig. 10. Insensitivity of lending activity to monetary policy. With two types of borrowers (good and bad) monetary policy may again have no effect on lending activity.

economy goes into a downturn, banks view their loan opportunity locus as shifting downward. In the absence of monetary authority intervention, loans would rapidly decrease. The monetary authorities can be viewed as leaning against a stream of water rushing the other way. They may be able to slow down the stream, but they have little chance of reversing its direction.

This does not fully answer the question, is there some rate of monetary expansion (some level of intervention by the monetary authorities) where credit would expand sufficiently to sustain the economy? Or more accurately, we should ask, is there some rate where the economy would ignite before inflation would ignite? To answer this, we need to return to the theme of monetary neutrality.

6.5. The classical dichotomy once again

Is it possible that no matter what the monetary authorities do, only the

\footnote{It should perhaps be emphasized that while the classical dichotomy is usually presented in the context of the standard neoclassical (Walrasian) model, it is more general: even with monopolic market structures, changes in the money supply should make no difference.}
price level would change, and nothing real would happen? As an aside, it should be noted that changes in the price level themselves (in an imperfectly indexed world) have real effects, as we have seen. The inflationary effects of monetary expansion would themselves represent positive redistributions towards firms, and thus serve to stimulate the economy.32

But the theory we have presented earlier provides the basis of a theory of nominal price rigidities33 as well: for so long as it is not common knowledge that money is neutral (and with imperfect indexing, the real redistributions associated with monetary expansions have real consequences), it would be risky for firms to act as if money was neutral. They surely will be uncertain whether increases in money have proportionate effects on the price level (for there is no empirical evidence supporting that belief); and given that uncertainty, they will not set their prices proportionate to changes in the money supply, at least in the short run. Non-neutrality is self-confirming (just as neutrality might be - and in most reasonable models, would be - self-confirming if all contracts were fully indexed and it were common knowledge that that was the case).34

Money, and more particularly, the banking institutions through which money is created in modern economies, are a central part of the capital market. It is, accordingly, not surprising that we should need to turn to a theory of the capital markets in general, and banks in particular, for an understanding of the role of money and credit, and the non-neutrality of money. But prices, in modern economies, are set by firms, not by mythical Walrasian auctioneers, who might indeed have been instructed to respond neutrally to changes in the money supply. Thus, we need simultaneously to construct a theory of bank behavior and firm pricing behavior: a task which we have now done.

7. Rationality and Keynesian economics

On the other side of the Atlantic, for a while, to invoke the name of Keynes was to associate oneself with the intellectual dinosaurs, or (to continue the metaphor) at least with a species (or at least a specious school) which, if not yet extinct, would surely soon be, and deservedly so. The ad

---

32There are other reasons that monetary policies, different rules concerning the rate of change of the money supply, would have real effects, even when the average rate of monetary expansion is fixed; for the pattern of returns to holding dollar denominated assets would change, and this would affect the demand for such assets relative to real assets.

33It is worth noting that several of the recently developed theories involving ‘convention equilibrium’ can also give rise to nominal rigidities; see, e.g. Stiglitz (1985, 1987b).

The ad hoc nature of the assumptions – their lack of basis in principles of maximization and rational behavior – condemned this line of thought to a perhaps not untimely death.

Today, most of us, I hope, recognize that the so-called ‘rational’ models were, in their own way, as ad hoc – and perhaps less realistic – than the Keynesian models: they were based on models of representative (identical) agents. Not only was that assumption unrealistic, results of Mantel, Debreu, and others had already shown that virtually any set of aggregate functions satisfying Walras’ Law could have been derived from some set of rational agents: once diversity was taken into account, the seemingly ad hoc assumptions most macro-economists had long employed could indeed be reconciled with maximizing behavior after all (and they were far more realistic than the particular parameterizations employed by many of those using the so-called rational approach, which employed utility functions, such as exponential or constant elasticity, the empirical implications of which were not only stringent, but clearly inconsistent with observed facts.)

There are other no less important elements of ad hocery that are frequently included in the so-called ‘rational’ models: monetary policies requiring cash in advance leave unexplained why credit cannot be used for facilitating transactions rather than cash. Indeed, this is an ad hocery for which no adequate explanation will ever be found – because it is in fact counterfactual, with most transactions being facilitated by credit, not cash. (It is not an explanation to construct a model in which individuals meet only once, so credit is not feasible: for that is no more true to life than the assumption that cash is required for all transactions. While it may be true that ad hocery, like beauty, may be in the eyes of the beholder, ad hocery is not the only vice from which a model may suffer: assumptions and conclusions which are counterfactual represent far worse sins.)

Our model invokes ‘rationality’ – indeed, more rationality than I believe is realistic. It differs from the neoclassical models not so much in that hypothesis as in the assumptions concerning information, and the related analysis of how markets and other economic institutions function with imperfect information. Consider, for instance, the standard criticism of Keynesian models, that they leave unexploited arbitrage opportunities: if, for instance, firms are off their supply curve, they could simply lower their prices. The imperfect information models upon which the analysis of this paper is based argue that there are no costless arbitrages: everyone is in fact

---

33 For instance, the exponential utility function, as convenient as it is for calculations, implies a zero wealth elasticity for risky assets.

35 While some new classical rational expectations models focus on consequences of information imperfections, they explore only a limited set of informational problems (e.g. concerning the money supply) – probably not even the central ones – and do not consider the possible consequences for how markets function.
maximizing, given his information, and given the costs of acquiring further information.

8. New and old Keynesian theories

Our model, while it bears a close resemblance to traditional Keynesian models, differs from it in several important respects, to which I should briefly call attention. [The differences are differences which I think Keynes would have liked. Hicks, who is to blame for many of the popular interpretations of Keynes, arrived at very similar views to those presented here in the years shortly before he died. See Hicks (1988).]

The first important difference was that Keynes failed to distinguish between the sources of finance, between debt and equity. This distinction plays an important role in our interpretation.

Secondly, and relatedly, Keynes had a very neoclassical theory of investment, with firms investing up to the point where the expected marginal return equalled the real interest rate. In this part of his analysis, he seemed to (uncharacteristically) ignore risk and credit market 'imperfections'. We have stressed both, explaining why risk cannot be divested, and why credit rationing may be prevalent. (The unimportance of the interest rate was one of the bases for Andrews' early criticism of Keynes.)

Having made these two 'errors', he was naturally led to the third: he focused on the interest rate as the mechanism by which monetary policy affected the economy. He focused on the 'money' side of banks' balance sheets, and how households' demand for money becomes extremely elastic in deep recessions. We have focused on the 'credit' side of banks' balance sheet. Money and credit are, of course, closely linked, but our analysis shifts the focus away from households' demand for money towards banks' supply of credit. In deep recessions, it is the unwillingness (and inability) of banks to make loans which obviates the effect of monetary policy, not the high elasticity of households' demand for money.

9. Policy prescriptions

Though there are these important distinctions between our theory and traditional Keynesian theory, there are also many similarities.

9.1. Channels of monetary policy

Elsewhere, we have shown how one can resurrect a modified IS–LM framework on more sound principles than those of Hicks: there is one important modification, which has some policy importance – there are some changes in monetary policy, like changes in reserve requirements, which
affect at any level of national income and any set of real and nominal interest rates, the availability of credit; and through this means, shifts in monetary policy may have a direct effect on the IS (the goods market equilibrium) curve, not mediated through changes in interest rates.

9.2. Wage and price flexibility

Keynes argued that wage–price flexibility would not restore the economy to full employment, because of adverse effects of wage decreases on aggregate demand. Wages and prices fell dramatically between 1929 and 1933, and this wage and price flexibility did not work: Keynes was right. One might argue that with sufficient wage and price flexibility the Walrasian equilibrium must necessarily be restored. (The real balance effect would eventually revive aggregate demand.) The Walrasian model, with its perfect information, perfect market assumptions is sufficiently flawed as to raise doubt about that conclusion: but what is clear is that estimated wealth elasticities of consumption, investment interest elasticities, etc. are such as to suggest that it might take an enormous fall in wages and prices to restore full employment. And our analysis suggests that so long as there are nominal credit contracts in place, the redistributions, the bankruptcies, and the resulting economic disruptions which would result would, at least in the short run, worsen matters, not make them better. It is not just the effect of real wage changes on the demand for consumer goods by credit-constrained households, but the effects on firms' real cash flows and balance sheets which are of crucial importance.

9.3. Ineffectiveness of monetary policy

The recent episode of monetary ineffectiveness in the United States – an ineffectiveness which we had predicted – provides further evidence in support of our theory. The Fed succeeded in substantially lowering interest rates on Treasury bills, but lending rates failed to fall commensurately, and lending activity failed to increase. Banks invoked both the arguments put forward in this paper as an explanation of their behavior: their net worth had been greatly reduced as a result of a series of 'misfortunes' – bad real estate loans, had oil and gas loans, and had Third World debt; and the high bankruptcy rate increased the perceived risk of lending, a risk which could only partially be mitigated through increased collateral requirements (which, by themselves, would have reduced the supply of qualifying borrowers).

While our theory thus conforms well with the policy perspectives of

37Neary and Stiglitz (1990) and Grandmont (1983) show that there may be adverse short run effects of price declines.
traditional Keynesian analysis — though the arguments underlying those conclusions are somewhat different — there are several respects in which our theory has policy predictions which differ from, or go beyond those of, traditional Keynesian theory. We briefly call attention to five.

9.4. Do government policies succeed in stabilizing the economy?

First, there has been concern that in the post-Keynesian era, volatility has remained high, perhaps as high as in the pre-Keynes era (though the latter contention has been hotly contested). Our analysis provides a partial explanation of this: assume the government succeeded in substantially reducing volatility. Then firms would undertake greater risk, e.g. by increasing their debt–equity ratio (viz. the markedly increased debt–equity ratios over the past decade). But then a variety of shocks would have an even greater impact on the economy. Thus, the net reduction in volatility may be much smaller: yet there may be substantial gains in economic efficiency, in allowing firms to undertake riskier (presumably higher return) activities than they otherwise would have.

9.5. Tax policy, risk sharing, and economic stability

Secondly, tax policies affect the efficiency of the economy's ability to bear risk. Domar and Musgrave long ago recognized that the government was like a silent partner. But the government is not like any partner that one would like to take on voluntarily: for while he shares profits, he does not fully share losses, and this lack of risk sharing may indeed exacerbate the consequences of limitations on the markets' ability to share risk, upon which we have already focused. The limitations on loss offsets are there for a good reason: to prevent tax abuses. The question is, are there ways in which the government can more effectively share risk, without at the same time opening the door to such abuses. I believe there are, but a fuller exploration of these must await another occasion.

9.6. Corporate taxation: average vs. marginal rates

Thirdly, many tax policies represent shifts in the tax burden, particularly in the current 'tight' fiscal climate of the United States and other countries. An attempt to stimulate consumption through individual tax reduction might be matched by an increase in corporate average tax rates. Believers in the standard Keynesian model might believe that such a shift would stimulate

---

38 See e.g. Romer (1987).
the economy. Consumers will spend more, and so long as firms' marginal
cost of capital or marginal returns are not affected, then investment will not
be adversely affected. Indeed, it is possible to increase average tax rates on
corporations and lower their marginal tax rates, and such a shift would
actually stimulate investment. Our analysis argues for the importance of
average corporate tax rates, for it is these which affect the firms' net worth,
and its ability and willingness to bear risks. And investment responses to
changes in net worth may indeed be quite large. Thus, the loss in aggregate
demand from the increase in the average corporate tax rate may more than
offset the gain in aggregate demand by consumers.

9.7. The importance of supply side responses

Fourthly, in many Eastern European countries, concern about inflation
has led to marked contractions in the availability of credit. Several commen-
tators [see, e.g. Calvo and Frankel (1991)] have noted that the resulting
economic contraction is dominated not so much by a reduction in demand
as by a contraction in supply, as firms without working capital have had to
cut back on their level of production. Old fashioned Keynesian models, with
their exclusive attention to aggregate demand, missed this potentially import-
ant aspect of the economy's response, an aspect which our analysis has
explicitly focused upon.

9.8. Small open economies

Fifthly, Keynesian analysis has always had a problem in small open
economies. On the one hand, in competitive versions, the country faces a
perfectly elastic demand for its products at the appropriate exchange rate: if
the country can sell anything, it can sell as much as it wants; there is no
problem of insufficient aggregate demand. On the other hand, with perfect
capital markets, real interest rates are determined internationally, so that
monetary authorities at most can affect the exchange rate. If investment is
affected, it cannot be because of effects on real interest rates, but only
because of effects on exchange rates.

Yet small open economies – including those exporting in competitive
markets – do experience economic fluctuations, and monetary policy often
seems effective and not just because of changes in exchange rates. Our theory
provides an explanation. The fluctuations in output are supply, not demand
driven. And our theory emphasizes that while markets for government T-bills
may work quite well, capital for loan purposes does not flow so freely.
Monetary policy can affect the credit institutions within a country, and by
that means, have a real effect on the economy. (Of course, even that effect
may get weakened, as multinational firms switch their borrowing from banks in one country to another.) Local banks, with localized knowledge, remain affected, and thus their lending activities, and the level of aggregate economic activity.\textsuperscript{39}

Concluding remarks

The last fifteen years have seen a revolution in our thinking about micro-economics, about how markets function. Principal-agent problems, incentive-compatibility constraints, moral hazard problems, selection problems, imperfect competition have become, if not household words, words familiar to all graduate students in economics. Even policy discussions are increasingly carried out using such terms – most notably, in the debates concerning the S & L and banking crisis in the United States.

Only gradually have we recognised that this change in our micro-economics necessitates a corresponding change in our macro-economic theories. In this paper, I have attempted to describe one part of that research program – that focusing on capital markets. It is an exciting research program – as many questions have been raised as have been answered. The theory is far from complete – though it is far more complete than it was but a few years ago. Other parts of the research program have involved the analysis of labor markets and the construction of simple, general equilibrium models [see, e.g. Greenwald and Stiglitz (1988c, 1990)] bringing together capital, labor, and product markets. High on the agenda for future research is further empirical verification of the theory – though, from our perspective, the fact that the theory is able to explain so many phenomena which, under the standard theory, seem so anomalous provides strong empirical support to the theory. At the same time, it is worth noting that many of the micro-economic aspects of the theory – the relative unimportance of equity, the adverse effect of equity issues on share prices, the dependence of investment on cash flow and firm balance sheet variables – have received increasing empirical support.\textsuperscript{40}

While, to be sure, the theory is far from complete, the results, both theoretical and empirical, obtained so far suggest that it holds out the promise not only of resolving some long standing macro-economic puzzles, but also of allowing us to design macro-economic policies which more effectively stabilize the economy.

\textsuperscript{39}A similar analysis suggests that banking regulations which affect the flow of funds within a country (such as laws restricting interstate banking) can have real effects; and reducing these barriers need not be welfare improving.

\textsuperscript{40}See, e.g. Hubbard (1990).
References


Bhattacharya, S., 1979, Imperfect information, dividend policy and the bird in the hand fallacy, Bell Journal of Economics 10, 259–270.


Myers, S. and M. Majluf, 1984, Corporate financing and investment decisions when firms have information that investors do not have, Journal of Financial Economics 13, 187-221.


Vale, B., 1989, Impact of central bank lending under asymmetric information in credit markets, Typescript (Stanford University, CA).