Why Is a Financial Crisis Important?
The Significance of the Relaxation of the Assumption of Perfect Competition

Yew-Kwang Ng*
Department of Economics, Monash University, Australia

Abstract

Under the usual assumption of perfect competition, we have money being neutral and changes in nominal aggregate demand cannot affect the real economic variables. If so, a financial crisis cannot be very important. However, the real world is characterized more by non-perfect competition when changes in nominal demand can affect real variables. This paper shows the important differences and explains the crux of these differences from both the demand and cost sides. It also provides a simplified general-equilibrium analysis of the economy and shows that, by concentrating on a representative firm and on how this firm is affected by macro variables and simplified interaction with other firms, macro analysis of the economy without assuming perfect competition is manageable with more realistic and richer results.

Key words: competition; financial crisis; market power; money neutrality; recession

JEL classification: E10; E30; D40

The 2008 financial crisis swept across the U.S. and the whole world in the last few months of 2008 (though in preparation much earlier) with a force never seen in the last eighty years. No one seems to doubt the devastating effects of a financial crisis of such force and extent. Not only have asset prices fallen sharply, no one doubts the big negative effects on the real economies it has caused and will continue to cause. On the other hand, few economists doubt the general acceptability and/or applicability of such basic economic assumptions and results as perfect competition and neutrality of money. However, as shown in Section 1 below, under such assumptions, a financial crisis would not really matter as it would not affect the real economy; it would not cause a recession and unemployment! However, the relaxation of the assumption of perfect competition alone changes this result, with a financial crisis now possibly very important, as shown in Section 2. Of course, there

Received August 4, 2009, revised September 28, 2009, accepted September 30, 2009.
*Correspondence to: Department of Economics, Monash University, Clayton, Australia 3800. E-mail: Kwang.ng@buseco.monash.edu.au. I am grateful to an anonymous referee for commenting on the first draft.
are other factors that may explain the importance of a financial crisis, such as time lags, nominal rigidities, etc. However, they are beyond the scope of this paper.

Although presented in the text mainly in the simple diagrams of classroom style, the analysis is really general equilibrium with interaction with other firms taken into account. (See the appendix for the full details.) Also, while the labour market is not explicitly modeled, its influence is incorporated in the analysis through the effects of aggregate output (which is related to the aggregate employment level) on the costs of firms.

I analysed the macroeconomic implications of non-perfect competition many years ago (Ng 1977, 1980, 1982, 1986). Despite the support of Marris (1991, p. 215) and Naish (1993), the significance has not been widely recognized and its importance to the role of a financial crisis not adequately analysed. Since then, there has been macro analysis taking into account of non-perfect competition, as surveyed by Dixon and Rankin (1994), the conclusion is that "Imperfect competition by itself does not create monetary non-neutrality… It is the combination of imperfect competition with some other distortions which generates the potential for real effects" (Dixon and Rankin 1994, p.178). In Section 2 below, it is shown that imperfect competition by itself may create non-neutrality. Also, in contrast to the specific functional forms used by other analysts, this paper uses graphical illustrations in the text and general utility, demand, and cost functions in the appendix. (On the graphical method, see also Naish 1993.) The use of general functions overcomes possible inconsistency with real-world empirics, as pointed out by Barro and Tenreyro (2006, P. 435) on some previous studies. Our conclusions are also consistent with the analytical and empirical result ‘that less competitive – or, at least, more concentrated – sectors feature more countercyclical movements in their relative output prices’ (Barro and Tenreyro 2006, p.435). There are also analyses based on the existence of frictions like adjustment costs (e.g. Ahlin and Shintani 2007, Akerlof and Yellen 1985, Naish 1988), and price-rigidity and disequilibrium analyses (see Benassy 2002 for a survey), in contrast to our equilibrium analysis without any informational imperfection, friction or rigidity. (For a recent empirical comparison of the sticky price and sticky information models, see Korenok 2008. For other possible links between the financial and the real sectors not considered here, see, e.g. Allen and Gale 2007, Bernanke and Gertler 1990, Holmstrom and Tirole 1997, and Kiyotaki and Moore 1997.)

The difference with the previous result and the crucial difference between perfect and non-perfect competition is discussed in Section 3, showing the relevance of both the demand and supply/cost sides. Section 4 explains the divergence in terms of the presence of an interfirm macroeconomic externality in the non-traditional cases. Section 5 concludes with analytical and policy implications and the appendix provides a mathematical model to back up the largely graphical analysis of the text. It should be emphasized that this paper deals only with the significance of the market power (or imperfection in competition) at the firm level, it does not discuss other relevant factors (such as asset price changes under nominal rigidity as in debts, herd psychology, time lags and money illusion) that may also make a financial crisis
of real significance. These factors are thus abstracted away (or regarded as exogenous and allowed as an exogenous change in aggregate demand or factors affecting the given response elasticities only) to focus on the point of concern here.

1. Unimportance of a Financial Crisis under Perfect Competition

Partly for simplicity and partly because the effects of a financial crisis typically takes place quickly, let us first focus on the short-run analysis defined by a given number of firms in the whole economy. (This is a common simplification in most macroeconomic analyses.) This allows the situation of the whole economy to be more directly reflected by that of a representative firm. The long-run analysis with a variable number of firms will also be briefly discussed later, without changing the crucial results substantially. (The representative-firm methodology may still be used in the long-run analysis, but the change in the number of firms must also be taken into account to supplement the picture of the firm itself.)

Under the usual assumption of perfect competition, a firm can sell as much of its product as it likes over the relevant range at the given market equilibrium price. In other words, it faces a horizontal demand curve for its product. The situation is illustrated in Figure 1. The firm is initially at a profit-maximization equilibrium point A where the upward-sloping marginal cost curve MC intersects the price line or demand curve (and also the marginal revenue curve) $d$, with price $p$ and output $q$. As the firm is representative, these are also the price and output levels (with appropriate scaling to take into account the number of firms in the economy) of the economy.

Now, let the nominal aggregate demand falls substantially as may be triggered by a financial crisis, a decrease in money supply, and/or some other factors. The lower demand shifts the market equilibrium price lower and hence shifts the demand curve faced by the firm from $d$ to $d'$. Ignoring frictions like temporary time lags (which will be reversed in time) and money illusion, the fall in the price level also shifts the marginal cost curve proportionately downward from MC to MC', as costs are also in nominal terms. These changes leave the firm continue to maximize profit at the new equilibrium point B that involves the same output level $q$. This gives the neutrality of money, i.e. a change (either an increase or a decrease) in money supply and nominal aggregate demand leads only to a change in the price level, not the real output and employment levels. The simple diagram of Figure 1 thus provides the microeconomic foundation for the much celebrated result of the neutrality of money. This result is derived in simple macro models based on the assumption that firms are perfectly competitive.

If changes in nominal aggregate demand do not affect real output and employment, a financial crisis cannot be very important. However, the neutrality result does not really apply in the real world, either in the short or long runs. What is wrong with the simple analysis above? Before answering this question in the next section, let us supplement the graphical analysis of Figure 1 with a simple
Consider a simple classical model with only one variable factor of production, with aggregate output \( Y \) being a function of the amount of labor \( L \) employed.

\[ Y = Y(L) \]  

(1)

Profit maximization under perfect competition involves the employment of labor until the marginal product of labor \( Y_L \) (where a subscript indicates partial differentiation) equals the wage rate, being the nominal wage rate \( W \) divided by the price level \( P \).

\[ Y_L = W/P \]  

(2)

The determination of the real wage rate \( W/P \) is influenced by the labor-supply condition reflected by the inverse labor-supply function \( f \):

\[ W/P = f(L) \]  

(3)

Equilibrium in the monetary sector is given by the classical demand-for-money function:

\[ PY = kM \]  

(4)

where \( M \) is the money supply and \( k \), a constant in this simple model, is the income velocity of money.
The two equations (2) and (3) determine the equilibrium values of real wage-rate \( \frac{W}{P} \) and employment \( L \). The latter in turn determines the real output level \( Y \) through the production function (1). Thus, all the real variables of the model (output, employment, and real wage-rate) are determined by equations (1) to (3), independent of (4). This implies that changes in money supply \( M \) must affect the price level \( P \) proportionately through (4), with \( k \) being constant and \( Y \) being determined independently by (1) to (3). The inability of money supply to affect the real variables is the neutrality of money. The independence between the real and the monetary sectors is the classical dichotomy.

2. Importance of a Financial Crisis under Imperfect Competition

As shown in Figure 1, a perfectly competitive firm at its equilibrium point A does not want to sell more of its product at the prevailing price, since doing so decreases its profits. If we ask one hundred firms in the real world whether they want to sell more of their products at the prevailing prices, at least over ninety of them will say yes; the more they sell, the more profits they make. The reason they do not sell more is that they cannot do so unless they decrease their prices. Thus, most firms in the real world faces downward-sloping demand curves for their products; they are not perfectly competitive. (This is partly due to product differentiation if for no other reasons than the difference in the location of firms.) The relaxation of the restrictive and unrealistic assumption of perfect competition makes a huge difference, even if all other simplifying assumptions such as no time lags and no money illusion are retained.

Even in the presence of non-perfect competition (inclusive of all market structures except perfect competition, with the demand curve for the firm’s product being downward sloping), the possibility of the neutrality of money and the associated unimportance of a financial crisis is not completely ruled out. This is illustrated in Figure 2 where the representative firm is shown to be facing a downward-sloping curve \( d \) (linearity in the demand and other curves are used only for ease of drawing but is not needed for the analysis) for its product and is initially at the profit-maximization equilibrium point A where its MC curve intersects its marginal-revenue curve MR. A decrease in nominal aggregate demand may shift the demand curve downward to \( d' \), with the new MR and MC curves intersecting at the same output level, with a lower price \( p' \). (As discussed below, such a shift will in fact be the case if the price level \( P \) is expected to decrease by the same proportion as nominal aggregate demand.) As the new profit-maximization price of the representative firm is in fact proportionately lower, this confirms the posited downward shift in the demand curve, making the new point B a consistent final equilibrium point. Thus, money may still be neutral.
However, with non-perfect competition, money neutrality or the impotency of changes in nominal aggregate demand in affecting the real variables is no longer necessarily true. Alternative outcomes become possible or even probable. (This is shown more rigorously in the appendix.) Consider Figure 3 where the representative firm is shown to be facing a downward-sloping curve $d$ for its product and is initially at the profit-maximization equilibrium point $A$ where its MC curve (not necessarily linear and horizontal) intersects its marginal-revenue curve MR. A decrease in nominal aggregate demand may shift its demand curve from $d$ to $d'$, its MR curve to MR'. (Such a shift will be the case if $P$ is expected to remain unchanged; see below.) The intersection of MR' with the new marginal cost curve (shown in Figure 3 to be the same as the old MC; this may not be needed, as an original MC, if downward-sloping, may shift downward to intersect MR' at the same point; or the MR may increase if the new demand curve becomes more price elastic; see the appendix for details) may give an unchanged profit-maximizing price but at a much lower output level $q'$. Moreover, if the original equilibrium point $A$ involves a long-run equilibrium zero-profit situation as depicted in Figure 3, the new short-run equilibrium point $B$ involves a big loss of the area DCBp. This is so because the AC curve is downward sloping, making a reduced output at an unchanged price profit-decreasing. This will force some firms to close down, making the unemployment situation (already brought about by the lower output) worse. This makes a financial crisis terrible. (See also the analysis of the long-run situation at the end of this section.)
For B to be really a final equilibrium point, it is important that it involves no change in price compared to A. This needs some explanation. Both demand curves \( d \) and \( d' \) are drawn given that the price level \( P \) is at a level corresponding to \( p \). (As the firm is representative, \( P \) should be the same as \( p \) at equilibrium, initial or final.) If the intersection of the new marginal revenue and cost curves is at a price slightly below B, with a slightly lower price and slightly higher output than B (such as B'), it may seem that this is just slightly better than B but is still a terrible situation. However, B' cannot be the final equilibrium. B' involves a lower price than \( p \). As this situation is representative, it means that most other firms also lower their prices, making the price level \( P \) lower as well. This lower \( P \) will lower the demand curve and make it more elastic than \( d' \) (more details of this shift in the next four paragraphs), and this will make the firm wanting to produce more and price less. But this leads to a further fall in \( P \) and the process iterates until the final equilibrium is achieved as the case of Figure 2 with output at \( q \) and a much lower price. (The full justification of this result is provided in the mathematical model in the appendix.)

How a change in \( P \) shifts the demand curve of the representative firm may be explained further. As derived in the appendix, the demand function for the product of the representative firm is given by

\[
q = f(p/P, \alpha/P)
\]
which specifies that the quantity demanded \( q \) is a function of the price of the firm \( p \) relative to the price level \( P \) (or the average price of all other firms) and real aggregate demand of the economy \( \alpha \) (\( = \alpha / P \); \( \alpha \) is the nominal aggregate demand). As \( q \) is a function of \( p \), \( P \) and \( \alpha \), the demand curve in a two dimensional diagram showing \( q \) as a function of \( p \) only is drawn for given levels of \( P \) and \( \alpha \). As either \( P \) and/or \( \alpha \) changes, the whole demand curve will shift in general. How do changes in \( P \) and/or \( \alpha \) shift the demand curve?

Consider Figure 4 where the demand curve \( d \) is drawn at some given levels of \( P \) and \( \alpha \). Consider a simultaneous decrease in \( \alpha \) and \( P \) by the same proportion. This leaves real aggregate demand unchanged. The firm should thus reckon that, if it also decreases its price \( p \) by the same proportion, quantity demanded will be unchanged, since both variables in (5) will then be unchanged. This is true at any point on the demand curve. Thus, a proportionate increase/decrease in \( \alpha \) and \( P \) shifts the demand curve upward/downward by the same proportion. The case of a decrease is shown as a shift from \( d \) to \( d' \) in Figure 4. This is also the case depicted in Figure 2.

Figure 4

Now consider a decrease in \( \alpha \) with \( P \) remaining unchanged. This decreases aggregate demand \( A \). The effect of this on a specific firm is uncertain, depending on such consideration as whether the good is an inferior good or not. However, for the representative firm, a fall in \( x\% \) in aggregate demand must decrease the quantity demanded also by \( x\% \) at unchanged prices, as otherwise the firm is not representative. Thus, the demand curve should shift leftward. However, in contrast to the case in the previous paragraph, we now have a real change, as \( A \) changes. This change may affect the demand elasticity. But this effect may go either way. For simplicity, Figure 4 shows the case where there are no such changes. (Such effects
are allowed in the appendix.) The demand curve thus shifts from \(d\) to \(d'\), similar to the case depicted (as \(d'\)) in Figure 3.

What about the effect of a change in \(P\) alone? It is difficult to figure out directly how such a change will shift the demand curve of the representative firm. However, we can get an answer indirectly. We already learned above that a proportionate decrease in both \(P\) and \(\alpha\) is to shift the demand curve vertically downward by the same proportion from \(d\) to \(d'\), as shown in Figure 4 and repeated in Figure 5. From this new situation, let us increase \(\alpha\) back to its original level, but leave \(P\) at the reduced level. This second change is just an increase in \(\alpha\) with \(P\) unchanged. We also know that such an increase in real aggregate demand will shift the demand curve proportionately rightward, from \(d'\) to \(d''\) as shown in Figure 5. Combining these two changes together, we have only \(P\) decreases with \(\alpha\) unchanged. So we know now that this is to shift the demand curve from \(d\) to \(d''\).

This lowers the demand curve and makes it more elastic in its upper section and raises it in its lower section. However, the lower section is not relevant as it involves negative marginal revenue. Thus, an increase/decrease in \(P\) alone is to raise/lower the demand curve and make it less/more elastic in the relevant range.

Figure 5

The discussion of how changes in \(P\) and \(\alpha\) affect the demand curve further support the different possibilities illustrated in Figure 2 and Figure 3 above, showing that the analysis is consistent with the macro and general-equilibrium repercussions, as explicitly provided in the appendix.

While the case of Figure 3 may be regarded as terrible, in fact, an even worse case may be possible, the case of cumulative contraction (even before considering the Keynesian income multiplier effect) illustrated in Figure 6. Here, the exogenous decrease in nominal aggregate demand does not only shift the demand curve leftward, but also makes it less elastic (as may be so if the remaining demand involves more essential items/units). The MC is also shown to be downward sloping...
and shifts downward as aggregate output decreases. (In the notations of the appendix, Figure 6 shows the case of $\eta^s < 0$, $\eta^f > 0$ and $D > 0$ and the resulting $(\eta^s + \eta^f - D)$ being negative.) For the case depicted, the new intermediate equilibrium point B involves a decrease in output and an increase in price. This change in $P$ makes B not yet the final equilibrium. This is so because the demand curves $d$ and $d'$ are both drawn given the initial price level $P$ being the one corresponding to $p$ at the initial point A. As discussed above, this increase in $P$ will further shift the demand curve for the product of the representative firm upward and make it more inelastic, leading to further reduction in output and increases in prices. The contraction will thus be cumulative as long as the condition for this remains unchanged. If this is the case, it makes a financial crisis even more terrible. This possibility may also partly explain the existence of stagflation.

The analysis so far is based on the simplifying assumption of a given number of firms. However, in a financial crisis that does trigger an economic crisis, some firms typically collapse, reducing the number of firms. Taking account of this requires a longer-run analysis that treats the number of firms $N$ as a variable. This is done in the appendix and shown not to alter the conclusions above substantially. In short, the different cases of money neutrality and non-neutrality, including cumulative contraction, are still applicable as in the short run, except that the conditions determining the various cases are made more complicated, as not only has the equality of MR and MC to be taken into account, but also the zero-profit condition (for long-run equilibrium) has to be accounted for. Thus, not only the shape and response of the marginal cost curve but also those of the average costs, as well as the effects of entries/exits have to be taken into consideration. For simplicity,
only the case corresponding to Figure 3 is illustrated in Figure 7 here. Readers interested in the more complete analysis are referred to the appendix.

As shown in Figure 7, from the initial equilibrium point A, a financial crisis that decreases nominal aggregate demand shifts the demand curve faced by the representative firm leftward, creating short-run losses (for cases such as illustrated in Figure 3; the case of purely nominal changes illustrated in Figure 2 is not considered here for the long run, on which see the appendix). This leads to the exit of some firms, lowering the degree of competition and making the new demand curve faced by the firm \(d'\) less elastic (than \(d\)). This lowers marginal revenue at given price, allowing the new equilibrium point B to involve lower marginal and average cost curves (through the effects of a smaller aggregate output \(Y\) in lowering costs) but unchanged price. This case involves a decrease in per-firm output plus a reduction in the number of firms. The long-run case of a cumulative contraction corresponding to the short-run case of Figure 6 may also be similarly illustrated.

In Figure 7, the MC curve is shown to shift downward (associated with a fall in input prices) as aggregate output decreases. This shift is however insufficient to avoid the reduction in aggregate output and employment, partly due to the large decrease in demand elasticity (exaggerated for clarity of graphical illustration). If the fall in input prices is higher than that associated in Figure 7, such as to shift the MC curve to a level lower than MC’, the intersection with MR’ will give a profit-maximization price lower than \(p\). If this is representative, the price level \(P\) will be lower. This will shift the demand curve of the firm to become more elastic (as explained above), eventually shifting the equilibrium back to \(q\). On the other hand, if the fall in input prices is no higher than that reflected by MC’, the economy may
be stuck at the low equilibrium point B for a long time. More remarkable, if we use the height of the MC curve of the representative firm to reflect the level of input prices, if input prices fall/respond more to the fall in aggregate output such that, at \( q' \), they fall to a level below that of MR', this allows the equilibrium to adjust back to \( q \) or A. After this equilibrium is established, the level of input prices in fact will be restored to a level consistent with the curve MC, much higher than if input prices do not adjust by enough and leave the economy being stuck at B.

3. The Crux of the Difference

As shown in the previous sections, just the relaxation of the assumption of perfect competition makes a big difference, changing the results of money neutrality and unimportance of a financial crisis. What is the crux of this big difference between perfect and non-perfect competition. There are two sides, the demand side and the supply or cost side.

On the demand side, under perfect competition, the demand curve for the product of the firm is horizontal. A horizontal demand curve cannot shift left and right (such a shift makes the curve remaining unchanged and hence is not really a shift); it can only shift up and down. However, an upward/downward shift means a change in price. For the representative firm, a change in its price means a change in the price level of the economy. In the absence of time lags and money illusion, this shifts the marginal cost curve by the same direction and proportion, leaving the intersection point of profit maximization unchanged. Hence, we have the irrelevance of changes in nominal aggregate demand as illustrated in Figure 1 in Section 1.

On the other hand, under non-perfect competition, the demand curve for the product of the firm is downward sloping. A horizontal demand curve may either shift (proportionately) up and down, as illustrated in Figure 2 in the previous section, leading, together with the appropriate condition in the supply or cost side, to no change in output, retaining the irrelevance of changes in nominal aggregate demand. However, it may also shift (proportionately) leftward and rightward, leading, together with the appropriate condition in the supply or cost side, to a change in real output with possibly no change in price, negating the money neutrality or the irrelevance of changes in nominal aggregate demand, as illustrated in Figure 3.

On the supply or cost side, under perfect competition, as the demand curve is horizontal, the marginal cost curve must be upward sloping to have a determinate equilibrium at the firm level. An upward-sloping MC curve means that marginal cost increases/decreases with an increase/decrease in output, making firms wanting to increase/decrease prices as output levels increase/decrease. This makes changes in nominal aggregate demand tend to be associated with price changes than with changes in real output.

On the other hand, under non-perfect competition, the demand curve for the product of the firm is downward sloping and the marginal-revenue curve is also typically downward sloping. (While an upward-sloping section in the MR curve cannot be ruled out, it tends to make the second-order condition unlikely to be
satisfied unless the MC curve is even more upward sloping.) This allows the MC curve to be either upward sloping, horizontal, or even downward sloping. There are also empirical and conceptual considerations suggesting that the MC curves of many firms may be largely horizontal over a large relevant range. Typically, a firm has a substantial amount of fixed costs and largely unchanged marginal costs of selling an additional unit over a large range. The non-upward-sloping nature of the MC curve makes the firm, as far as the cost-side consideration is concerned, not to change its price as its output changes. This makes the cases of non-neutrality of Figures 3 and 6 more likely to prevail.

In terms of the simple classical model discussed in Section 1, once the assumption of perfect competition is relaxed, we no longer have (2) which should be replaced by

\[ \mu Y_t = W \]  

(2')

or the equality of the marginal revenue (\( \mu \)) product with the nominal wage rate. This simple replacement destroys the classical dichotomy. The first three equations (1), (2') and (3) no longer definitely determine the three real variables \( Y, L, \) and \( W/P \). Instead of one real variable \( W/P \), we have two nominal variables \( \mu \) and \( W \). Moreover, a change in money supply \( M \) may affect \( \mu \). For the case of Figure 3, a reduction in nominal aggregate demand shifts the demand curve from \( d \) to \( d' \), lowering \( \mu \) at a given output level, destroying the classical dichotomy between the monetary and the real sectors.

Contrary to the analysis above, economists who have included non-perfect competition in their macro analysis concluded that "Imperfect competition by itself does not create monetary non-neutrality… It is the combination of imperfect competition with some other distortion which generates the potential for real effects" (Dixon and Rankin 1994, p.178). In fact, they have only shown that a real equilibrium under imperfect competition can still be an equilibrium even if the money supply changes. However, they have not shown that a change in the money supply (or any other factor causing a change in nominal aggregate demand) may not trigger a shift from one real equilibrium to another. Such a shift may not be relevant if the equilibrium is unique. However, the presence of non-perfect competition may make a model with a unique equilibrium into one with multiple or even a continuum of equilibria and hence make the above-mentioned shift relevant, making money possibly non-neutral even without any friction, as shown in Section 2 above and in the appendix.

4. The Interfirm Macroeconomic Externality

The difference between perfect and non-perfect competition may also be seen by considering the interfirm macroeconomic externality. Here, for the non-traditional cases (calling the money neutrality case the traditional one), the expansion/contraction of one firm benefits/harms others before the Keynesian
income-multiplier effect (which amplifies the externality) operates. For the short-run model presented in the appendix, we may examine the effects of an expansion/contraction of one representative firm on the short-run equilibrium real profits of another representative firm. We have

\[
\frac{\partial R_i}{\partial q_j} = \frac{1}{N} \left[ \frac{1}{\eta} \left( 1 - \eta^\omega \right) + 1 - \eta^{cr} - \eta^\omega \right] - \eta^{cr}
\]

(6)

where \(q_j\) is the output level of firm \(j\) (assumed representative) and \(R_i\) is the real profits of another firm \(i\) (also representative), \(N\) is the number of firms in the whole economy, \(\eta\) is the price elasticity of demand of the representative firm, \(\eta^\omega = (\partial a/\partial b) b/a\) for any \(a\) and \(b\) is the response elasticity of \(a\) with respect to \(b\) (or the endogenous effect of a change in \(b\) on \(a\) in proportionate terms), \(\alpha\) is the nominal aggregate demand in the economy, \(C\) is the total cost of the firm, \(P\) is the price level and \(Y\) is the aggregate output of the economy.\(^3\)

First, we may note that, if \(\eta^{cr}\) is positive and large, an expansion/contraction of one firm may reduce/increase the profits of another firm by shifting the cost curve of the latter. This is true even under perfect competition for the short run before the price adjusts to restore zero profit. For the long-run case, the effect is necessarily zero due to the zero-profit requirement. However, a positive and large value for \(\eta^{cr}\) is unfavourable for the non-traditional cases to be applicable. To concentrate on the non-traditional cases, let us ignore \(\eta^{cr}\) or let it equal zero. The role of \(\eta^\omega\) is to capture the Keynesian income-multiplier effect that reinforces the externality. To concentrate on the externality effect as such, ignore that as well. The value of \(C\eta^\omega - 1\) captures such effects as time lags and money illusion that makes \(\eta^\omega\) less than one, i.e. making costs respond less than proportionately to \(P\), reinforcing the externality effect. Ignore that too. Then, we are left with

\[
\frac{\partial R_i}{\partial q_j} = \frac{1}{N} \left[ \frac{1}{\eta} \left( 1 - \eta^\omega \right) \right]
\]

(6')

First note that, for the case of perfect competition, the right hand side of (6') vanishes as whatever value on the numerator is divided by infinity twice (note that \(\eta\), being the price elasticity of demand, equals minus infinity under perfect competition). It may be noted that our model is general such that whenever we assume perfect competition, we get the traditional result derived under perfect competition. However, under non-perfect competition, \(\eta^2\) is positive and finite. Since \(N\) is positive and \(\eta^\omega\) is positive but less than one, we have the right hand side of (6') being positive, i.e. an increase/decrease in output by a representative firm increases/decreases the real profits of another representative firm in the non-traditional cases.

The fact that the positivity of this interfirm macroeconomic external effects depends on the shortfall of \(\eta^\omega\) from one may also be explained. Given a downward-sloping demand curve, an attempt to produce/sell more involves a lower
price, a lower $P$ decreases nominal aggregate demand $\alpha$, but by less than proportionately. Thus, real aggregate demand $A = \alpha / P$ increases. This shifts the demand curves of other firms rightward and hence increases their profits; as their AC curves are downward sloping, the ability to sell more products at given prices increases their profits.

While the right hand side of (6') is positive, it is small as it has $N$ in its denominator and $N$, being the number of firms in the whole economy, is a large number. However, (6') only specifies the effect of an expansion by one firm on another firm. For the effects on all firms, we have to multiply by $N$. If all firms expand simultaneously, the effects of this on all firms have to be multiplied by another $N$. This double multiplication may then turn the right hand side of (6') from a small value into a large value. The per-firm value of this could be in the order of the area DCBp in Figure 3, certainly a large proportion of the total output of the firm.

We may also understand the interfirm macroeconomic externality by comparing the different situations (following the same decrease in nominal aggregate demand) illustrated in Figure 2 and Figure 3. In the case of Figure 3, firms keep their prices unchanged and produce less in the face of reduced demand. In the case of Figure 2, firms reduce prices and keep output unchanged. The equilibrium point B in Figure 3 involves a lower output and a higher price compared to that in Figure 2. In the face of a decrease in nominal aggregate demand, if all firms lower prices (and factor prices are also correspondingly lowered and hence reducing costs) to avoid having to decrease output, they could all maintain output as in the case of Figure 2. The attempt by a firm to decrease its price and avoid a reduction in output produces external benefits on other firms. However, unless all firms do that, a single firm is powerless in safeguarding the higher equilibrium output $q$. In the situation of Figure 3, if a firm tries to increase output from the point B by reducing its price, it will increase its losses as its MR will then be less than its MC. But if all firms expand from $q'$ or avoid reducing output from $q$ to begin with to achieve the situation of Figure 2, they may all be able to avoid the huge per-firm loss of DCBp.

5. Concluding Remarks

Few if any firm in the real world is perfectly competitive. However, the assumption of perfect competition may facilitate certain analysis such as the fully general equilibrium analysis that throws light on the working of the economy. Nevertheless, where the assumption is not necessary, it is much better to replace it with more realistic ones. Our analysis, especially the mathematical appendix, shows that the analysis of the macro economy need not be based on the misleading assumption of perfect competition. We have seen that the relaxation of this assumption alone destroys the classical dichotomy between the real and the monetary sectors, makes money possibly non-neutral, and makes a financial crisis possibly of very important consequences for the real economy. Though not as simple as the case of perfect competition, by concentrating on the microeconomics
of the representative firm but taking account of the role of macro variables like aggregate demand, the price level, and aggregate output, and considering the general-equilibrium interaction effects in a simplified way, our analysis is manageable. Perhaps it is time that macro analysis should move towards more realistic models.

Of course, our analysis tackles mainly the relaxation of the perfect competition assumption; there are many other factors that may be relevant in making a financial crisis important, such as the role of asset/wealth, time lags, herd psychology, etc. There are also many important economic issues beyond the financial crisis. However, these are beyond the scope of the present paper.

Though only relaxing the assumption of perfect competition, our analysis already provides some insights relevant not only for conceptual analysis but also for practical policy. It shows that in an economy with non-perfect competition, changes in nominal aggregate demand, as may be triggered by the loss in consumer and business confidence due to a financial crisis, may lead to large and even cumulative falls in real output and employment. The attempts of many central banks to safeguard liquidity in the current crisis can thus be amply justified. Similarly, the attempts by many governments to reduce taxes and increase government spending may also be justified.

Our analysis also suggests that, to prevent the economy from being stuck at a low-level equilibrium point, it is important to have input prices adjust sufficiently to changes in aggregate output. Temporary falls in input prices may restore the economy back to a high equilibrium at which the level of input prices will be much higher than the level at the low equilibrium point. Another point is the importance of avoiding firm bankruptcies. This is so not only because of the more observable effects of reducing output and employment but also because of the unfavourable effects on competition (making the price elasticities of demand faced by firms lower) that may lead to higher prices and lower output levels.

Appendix: Mathematical Analysis

This appendix supplements the largely graphical analysis of the text with a mathematical model of simplified general equilibrium.

For the general case of either perfect or non-perfect competition, we may regard the product of each firm as a distinct good; then, from the maximization of a general utility function

$$U' = U'(x_1, x_2, \ldots, x_N)$$  \hspace{1cm} (A1)

where $U'$ = utility of individual $i$, $x_i$ = amount of the $g$th good consumed by individual $i$. The maximization of utility specified in (A1) above subject to a budget constraint
\[ \sum_{g=1}^{G} p_g x'_{ig} = \alpha' \]  

(A2)

where \( p_g \) = price of the \( g \)th good, \( \alpha' \) = budget constraint = nominal aggregate demand of individual \( i \), we may derive the following demand functions

\[ x'_g = x'_g(p_g, p_z, ..., p_N, \alpha'), \quad i = 1, ..., I; \quad g = 1, ..., N \]  

(A3)

where \( I \) is the number of consumers. Summing over all consumers, we have

\[ q_s = \sum_{i=1}^{I} x'_g = F_s(p_g, ..., p_N, \alpha'_i, ..., \alpha'_I) \]  

(A4)

Now, to avoid a fully general equilibrium analysis, two simplifications are adopted. The first is one used by virtually all macro and aggregative studies of ignoring distributional effects by replacing the vector \( \alpha'_i, ..., \alpha'_I \) by \( \alpha = \sum \alpha' \).

The second simplification is to concentrate on the representative firm and replace the price vector of all other firms by the average price. Without loss of generality, call this representative firm "Firm 1". From (A4), we then have

\[ q_i = f_i(p_1, P) \]  

(A5)

where \( p_1, ..., p_N \) has been replaced by \( P \), the average price of \( p_1, ..., p_N \). Due both to the fact that Firm 1 is representative, and the fact that it is assumed small relative to the whole economy, this average price \( P \) is also the general price level of the economy. A fully general equilibrium analysis (Ng 1986, App.3I) has been used to show that (1) for any (exogenous) change (in cost or demand) there exists (in a hypothetical sense) a representative firm whose response to the change accurately (no approximation needed) represents the response of the whole economy in aggregate output and average price, and (2) a representative firm defined by a simple method (that of a weighted average) can be used as a good approximation of the response of the whole economy to any economy-wide change in demand and/or costs that does not result in drastic inter-firm changes. These fully general-equilibrium results provide a strong methodological justification of our representative-firm approach.

Since demand functions are homogeneous of degree zero in all prices and the budget, we may divide all elements in the functional form in (A5) by \( P \) (i.e. multiply all elements by \( 1/P \)) to obtain, dropping subscripts,

\[ q = f(p/P, \alpha/P) \]  

(A6)

where the effect of \( P/P \), being a constant, is defined into the functional form of \( f \).

The representative firm is assumed to take the aggregate variables as given and maximize its profit with respect to its own output or price. Its profit function is
\[ pf\left(p/P, \alpha/P\right) - C(q, Y, P, \epsilon) \tag{A7} \]

where \( C \) = total cost, \( Y \) = aggregate output of the economy, \( \epsilon \) = exogenous factors affect costs. The possible effects of \( Y \) on \( C \) may work through the input market. [If desired, one may use \( w(P, Y) \), where \( w \) is the (vector of) input prices in place of \( P \) and \( Y \) in the cost function in A7 above.] It may be noted that the cost function is rather general. The first-order condition for the maximization of (A7) gives

\[
\mu = p\left(1 + \frac{1}{\eta}\frac{\alpha}{\frac{P}{P'}}\right) = c(q, Y, P, \epsilon') \tag{A8} \]

where \( \eta = \left(\frac{\partial q}{\partial P}\right)p/q, c = \partial C/\partial q \) are respectively the price elasticity of demand and marginal cost and \( \mu \) is marginal revenue. This equation simply specifies the equality of MR with MC.

From the representativeness of the firm and the requirement of equilibrium, we have

\[
P = p \tag{A9} \]
\[
\alpha/P = A = Y = qN \tag{A10} \]

where \( A \) = real aggregate demand, \( N \) = number of firms. (The latter is taken as given in the short-run exercise. Otherwise, an additional equation of free entry/exit may be added, as done below. The possible feedback of changes in profit is allowed through the effect of real income on aggregate demand in Eq. A11 below, abstracting from distributional effects.)

The nominal aggregate demand of the economy is taken to be a function of \( P \), \( Y \) (real income which equals real output at equilibrium) and some exogenous factors \( \epsilon'' \) which should include the money supply.

\[
\alpha = \alpha(P, Y, \epsilon''); \quad 1 > \eta'' > 0; \quad 1 > \eta'' > 0 \tag{A11} \]

where the restriction \( 1 > \eta'' > 0 \) is similar to the case of the Keynesian cross diagram that the slope of \( C + I \) is positive but less than one to avoid an explosive system. Similarly for \( (1 - \eta''), \) and \( \epsilon'' \) is the (vector of) exogenous factors affecting nominal aggregate demand. (A11) is a very general function and include the simple Keynesian and Monetarist aggregate demand functions as special cases. This completes the specification of our very simple, general, but powerful model. We turn now to the comparative-statics analysis. The total differentiation of (A8), after substituting in the total differentiation of (A9) and (A10), division through by marginal revenue \( \mu \) or marginal cost \( c \), gives
\[
\left(1 - \eta^{e}\right)\frac{dP}{P} - \left(\eta^{e} + \eta^{c} - D\right)\frac{dY}{Y} = \frac{d\alpha}{c}
\]  \hspace{1cm} (A12)

where \( \eta^{e} = (\partial c/\partial y)(y/x) \), \( D = (\partial \mu/\partial A)(A/\mu)p \), \( P = -(p/\eta)(\partial \eta/\partial A)(A/\eta) \) is the proportionate effect of real aggregate demand on marginal revenue at given prices through possible effects on the price elasticity of demand, \( d\alpha = (\partial c/\partial \alpha) \) is the exogenous change in marginal cost.

The total differentiation of (A11), after dividing through by \( \alpha \) and substituting in \( d\alpha/\alpha = dP/P + dY/Y \) from the total differentiation of (A10), gives

\[
\left(1 - \eta^{e}\right)\frac{dP}{P} + \left(1 - \eta^{c}\right)\frac{dY}{Y} = \frac{d\alpha}{\alpha}
\]  \hspace{1cm} (A13)

where \( d\alpha = (\partial \alpha/\partial \alpha) \) is the exogenous change in nominal aggregate demand.

Substituting \( dY/Y \) and \( dP/P \) from (A13) in turn into (A12), we obtain (A14) and (A15).

\[
\Delta \frac{dP}{P} = \left(\eta^{e} + \eta^{c} - D\right)\frac{d\alpha}{\alpha} + \left(1 - \eta^{c}\right)\frac{d\sigma}{c}
\]  \hspace{1cm} (A14)

\[
\Delta \frac{dY}{Y} = \left(1 - \eta^{c}\right)\frac{d\alpha}{\alpha} - \left(1 - \eta^{c}\right)\frac{d\sigma}{c}
\]  \hspace{1cm} (A15)

where \( \Delta = (1 - \eta^{e})(1 - \eta^{c}) + (1 - \eta^{c})(\eta^{c} + \eta^{c} - D) \), \( c = \partial C/\partial q = \) marginal cost of the representative firm, \( C = \) total cost, \( \eta^{e} = (\partial a/\partial b)b/a \) for any \( a \) and \( b \), e.g. \( \eta^{c} = (\partial c/\partial y)y/c \) is the (proportionate) change in MC as the firm itself changes its output \( q \) (the sign and magnitude of \( \eta^{e} \) thus relates to the slope of the MC curve), \( \eta^{c} = (\partial c/\partial y)y/c \) is the (proportionate) shift in MC as the output \( Y \) of the whole economy changes, \( \eta^{c} = (\partial a/\partial y)y/\alpha \) is the (proportionate) change in nominal aggregate demand \( \alpha \) in response to an increase in aggregate output \( Y \) (which equals aggregate income in equilibrium) and is similar to the Keynesian marginal propensity to spend. Also, a total differentiation sign \( d \) before a variable with a bar indicated the exogenous change in that variable; \( D \) is the (proportionate) effect of real aggregate demand on the marginal revenue of the representative firm at given prices, through changing the price elasticity of demand. It may also be noted that, for stability and for signing the comparative statics, \( \Delta \) may be taken as positive. (This does not rule out the possibility that \( \Delta \) may be zero or even negative at a particular range, causing a continuum of equilibria and instability. However, the negativity/instability must be reversed beyond this range where a new equilibrium may be established. Similarly, in the case of the Keynesian cross diagram, while the C+I curve may be steeper than 45 degrees at some location, it must become less steep for a final equilibrium to be established.)

The two equations above indicates that the effects on the price level and aggregate output depends on the exogenous changes in demand and costs as well as the endogenous response variables, including the slope of the MC curve (of the representative firm), how much MC responds to aggregate output and the price level...
(shifts in the MC curve), how much the price elasticity of demand changes in response to real aggregate demand, how much the nominal aggregate demand changes in response to real aggregate income and the price level. An estimate of these changes and responses would then give us our estimate of the effects on the price level and aggregate output using the above equations.

These two equations are the basic comparative-statics results which can be used to analyze the general equilibrium effects of internal and/or external economy-wide changes in demand and costs on the price level and aggregate output of the economy.

In particular, to analyze the effects of an exogenous change in nominal aggregate demand as may be relevant for a financial crisis, we may abstract away any exogenous changes in costs by taking \( \alpha D/c \) in (A14) and (A15) as zero. (This is not a partial-equilibrium analysis, as endogenous changes in costs are allowed. Rather, it is just ‘one thing at a time’.) We then have, from (A14) and (A15),

\[
\frac{\Delta P}{\Delta \alpha} = \frac{(\eta^p + \eta^e - D)}{\Delta} \quad (A16)
\]

\[
\frac{\Delta Y}{\Delta \alpha} = \frac{(1 - \eta^p)}{\Delta} \quad (A17)
\]

These two equations specify how a change in nominal aggregate demand \( \alpha \) exogenously may affect the price level \( P \) and aggregate output \( Y \) proportionately.

Let us first consider the case of perfect competition. Here, we have \( \eta^p \) positive as the MC curve must be upward sloping as explained in Section 3 in the text. Next, \( \eta^e \) is non-negative if we ignore (as invariably done so in macro analysis) such factors as Marshallian external economies, as then an increase in aggregate output does not decrease the marginal costs of firms. \( D \), the change in the marginal revenue of the firm as aggregate demand changes at unchanged prices through possible changes in the price elasticity of demand (for the product of the firm), is necessarily zero under perfect competition since the price elasticity of demand is (minus) infinite and remain infinite and hence does not change. This makes \( \eta^e (\eta^p + \eta^e - D) \) necessarily positive. Also, ignoring such frictions as time lags and money illusion, marginal costs adjust proportionately to prices, making \( (1 - \eta^p) \) zero. Thus, from (A16) and (A17), we have a change in nominal aggregate demand exogenously changes only the price level without affecting the output level.

If we do not include the endogenous change in nominal aggregate demand \( \alpha \) (i.e. taking \( \eta^\alpha \) and \( \eta^\alpha \) in \( \Delta \) as equal to zero), we have, from (A16) and (A17), the price level \( P \) decreases/increases proportionately with the exogenous decrease/increase in nominal aggregate demand. Including the endogenous changes, we have

\[
\frac{\Delta P}{\Delta \alpha} = \frac{1}{1 - \eta^\alpha} > 1 \quad (A16')
\]
as $\eta^\omega$ is positive but smaller than one from (A11). Thus, the endogenous change reinforces the exogenous change. This reinforcement effect through $\eta^\omega$ may be called the price-multiplier effect, similar to the Keynesian income-multiplier effect.

An exogenous decrease in $\epsilon^\omega$ (due to say a decrease in money supply or loss in consumer and investor confidence in a financial crisis) decreases $\alpha$ by say 10% at the initial income and price levels (see eq. A11). This first leads to a decrease in the price level by 10% in the present case of perfect competition. This decrease in $P$ further decreases nominal aggregate demand $\alpha$ and leads to a further decrease in $P$. If $\eta^\omega = 1/2$, the price multiplier equals 2 and the total decrease in $P$ is twice as large as the initial exogenous decrease in $\alpha$. In the present case with no effect on $Y$, the endogenous effects through $Y$ do not take place. For an increase (in contrast to that of a decrease mainly being concerned here) in nominal aggregate demand, a large price multiplier may partly explain some inflationary episodes such as the oil price shocks. Our results are also consistent with the finding by Rotemberg and Woodford (1996) that a small change is the price of a factor of production of relatively small proportionate employment may lead to large effects on prices and output.

Now consider the more realistic and general case of non-perfect competition. Even in this case, we may still have $(\eta^\omega + \eta^\gamma - D)$ being positive and $(1 - \eta^\gamma)$ being zero, making the results similar to the case of perfect competition, as shown in Figure 2 in the text. However, with non-perfect competition, $(\eta^\omega + \eta^\gamma - D)$ may no longer be positive. In particular, $\eta^\omega$, $\eta^\gamma$ and $D$ may all be zero, which is the case illustrated in Figure 3 in the text. More generally, the values of these three variables or response parameters may combine to give a zero or even negative value for $(\eta^\omega + \eta^\gamma - D)$. For the case of a zero value, we have an exogenous decrease in nominal aggregate demand (as may happen in a financial crisis) triggering a decrease in real output (and hence employment) with no effect on the price level, similar to the special case illustrated in Figure 3. For the case of a negative value, we have the case of a cumulative contraction (or cumulative expansion for an increase in aggregate demand) illustrated in Figure 6 in the text. In either case, a financial crisis may lead to a big fall in real output and employment and cause an economic crisis as in the case of 1929, if not reversed by appropriate counter-recession policies.

It may be noted that a negative $(\eta^\omega + \eta^\gamma - D)$ must not be mechanically interpreted as a fall in aggregate demand leading to a decrease in prices and an increase in output. This interpretation is a similar mistake as, for the case of a steeper-than-one $C+I$ curve in the simple Keynesian income determination diagram, taking an increase/decrease in exogenous investment as leading to a decrease/increase in income/output. A $C+I$ curve steeper than the 45 degree line is also one of cumulative expansion/contraction.

For the long run where the number of firms is allowed to change, we allow $N$, the number of firms, to enter the demand function and also add the additional condition of zero long-run profit

$$pf(p|p,\alpha|\theta)-C(q,Y,P,\epsilon^\gamma) = 0$$  \hspace{1cm} (A18)
It may be thought that, the zero-profit condition should only be applied to the marginal firm, not to the representative firm. However, under the condition of free entry and competition in the factor market, whatever factors that account for a positive profit of any firm will have their prices bid up to make the genuine economic (though not accounting) profit zero. With this additional condition, similarly to the derivation of (A14) and (A15) above, we may derive the comparative results for the long run, yielding:

\[
\nabla dP/P = [M\eta^{ct} + (E + \eta^\alpha)\eta^{ct} - M(D + E)](d\overline{c}/\alpha) + (1 - \eta^{ct})(E + \eta^\alpha)(d\overline{C}/C) + (1 - \eta^{ct})M(d\overline{c}/c)
\]

\[
\nabla dY/Y = [M(1 - \eta^{ct}) + (E + \eta^\alpha)(1 - \eta^{ct})](d\overline{c}/\alpha) - (1 - \eta^{ct})(E + \eta^\alpha)(d\overline{C}/C) - (1 - \eta^{ct})M(d\overline{c}/c)
\]

where

\[
\nabla = (1 - \eta^{ct})[M\eta^{ct} + (E + \eta^\alpha)\eta^{ct} - M(D + E)](d\overline{c}/\alpha) + (1 - \eta^{ct})[M(1 - \eta^{ct}) + (E + \eta^\alpha)(1 - \eta^{ct})]
\]

\[M = (p - c)/\rho\] is the markup of price over marginal cost, \(E = (\partial\mu/\partial N)\mu\) at given prices (where \(\mu =\) marginal revenue of the representative firm) is the effect on MR of the entry of new firms through a higher absolute price elasticity of demand due to increased competition.

The long-run effects of a shock may then be analyzed using (A19) and (A20) above by estimating the long-run exogenous changes in aggregate demand and costs as well as the long-run endogenous response variables contained in these two equations. Comparing (A19) and (A20) with (A14) and (A15), we may see that all cases discussed for the short run above remain relevant except that conditions determining the various cases become more complicated, with \([M\eta^{ct} + (E + \eta^\alpha)\eta^{ct} - M(D + E)]\) replacing \((\eta^{ct} + \eta^\alpha - D)\) and with \([M(1 - \eta^{ct}) + (E + \eta^\alpha)(1 - \eta^{ct})]\) replacing \((1 - \eta^{ct})\) as the relevant values determining the various cases. This is so since, for the long run, the response of the average or total costs and the effects of entry/exit must also be taken into account.

In particular, the case with a real output adjustment and no price changes following an exogenous changes in nominal aggregate demand (as illustrated in Figure 3 for a special short-run case and in Figure 7 for a long-run case) is still possible if \([M\eta^{ct} + (E + \eta^\alpha)\eta^{ct} - M(D + E)]\) is zero and the cumulative contraction/expansion case is relevant if this expression is negative. Comparing the likelihood of these non-traditional cases (i.e. calling the money neutrality case as traditional) between the short and long runs, we may note that while the effects of a change in aggregate demand is more likely to have a significant and negative effects on costs, making the non-traditional cases less likely to prevail in the long run, the entry/exit effect \((E\) is positive) makes the non-traditional cases more likely to prevail in the long run.
Notes

1. Despite the passing of well over a decade, this situation has remained unchanged; see, e.g. Dixon (2007).
2. In some more complicated models, the two may not always go together.
3. The derivation of (5) is available from the author.
4. This appendix was developed from the earlier analysis in Ng (1980) but starting from the more basic level of utility maximization and relaxing some simplifying assumptions. Apart from analyzing the effects of a financial crisis, the text also adds on to the earlier analysis by focusing more on the crux of the differences between perfect and non-perfect competition.
5. If \( (\eta^* + \eta^* - D) \) and \( (1 - \eta^*) \) are both zero, our equations show indeterminate results. Further graphical analysis and common sense then suggest that the results then depend on expectation which will then be self-fulfilling.

References


