

**THE UNIT ELASTICITY HYPOTHESIS  
AND THE INDETERMINACY OF OUTPUT**

Robert E. Hall

Department of Economics and Hoover Institution  
Stanford University  
National Bureau of Economic Research

March 4, 1986

This research is part of the NBER's Program on Economic Fluctuations. I am grateful to Michael Knetter for outstanding assistance and to numerous colleagues for helpful discussions and advice.

## *Introduction*

Before Keynes, macroeconomists believed that aggregate output grew along a steady growth path with occasional brief interruptions. In the post Keynesian synthesis that has ruled macroeconomics since the 1960s, the story is different only in that the departures from full employment may last several years. But the consensus holds that the price-wage adjustment process gradually returns the economy to its full-employment equilibrium. The unique resting point of a modern model is full employment. Virtually all of the criticism of this consensus has come from a school that sees the equilibrating process as even faster than the one contemplated in the consensus. Pure equilibrium models, where adjustment is instantaneous, have even attracted a few followers.

Recently, the consensus has started to weaken in a startling new way. First, Oliver Hart (1982) and a number of others have shown that a full-employment equilibrium is not invariably the outcome in an economy with less than perfectly competitive markets. The drive to full employment implicit in the consensus generally rests on competitive forces. Prices continue to fall as long as supply exceeds demand, because individual firms see an opportunity to make added profit by selling a new unit of output at a price that exceeds its marginal cost of production. The only mystery is why it takes so much time for this realization to be translated into lower prices and higher output. By contrast, in an imperfectly competitive market, a seller is concerned with whether marginal revenue exceeds marginal cost. The essence of the new theoretical literature is to point

out that the equality of marginal revenue and marginal cost can occur at less than full employment. In competition, the equality of price and marginal cost always takes place at full employment.

The finding of a sharp distinction between competition and imperfect competition as a matter of theory is complemented by empirical work suggesting that most markets in the U.S. economy are far from competitive. These findings are summarized by Bresnahan (1986) and supported by additional work in a different framework in Hall (1986).

The other new development is the finding that the stochastic behavior of output is inconsistent with a tendency to return to a full-employment trend path. The earliest mention of this point seems to be Nelson and Plosser (1982); a number of other authors have studied the point since then. It now seems reasonably well established that output in all modern capitalist economies has a dominant random-walk component. When output changes by one percent, it changes the optimal prediction of output for the indefinite future by about the same one percent. Such behavior flatly contradicts the view that most stochastic shocks are eventually worked out through the price adjustment process.

This paper presents and discusses a hypothesis that explains the absence of effective market-clearing forces in the aggregate and in individual markets. The hypothesis is compatible with the random-walk character of output and with a good deal of other evidence that is puzzling and even baffling in competitive analysis. The hypothesis involves imperfect competition, but it is not implied by Hart's model nor by its successors, even the ones that involve indeterminacies of equilibrium.

I call it the *unit elasticity hypothesis*. The condition for profit maximization can be expressed as the requirement that the elasticity of

demand with respect to the part of the price captured by fixed factors (price less marginal cost) be equal to one. In received theory, this condition is satisfied at the point of maximum profit; this point is considered unique, so the condition fails at other prices and corresponding levels of output. Under the UEH, the condition holds over a range of prices. The firm is indifferent among all the prices and levels of output for which the UEH holds.

The paper argues first that the available evidence, both on the stochastic process of output and other types of evidence, is consistent with the UEH and not with other views. The evidence includes data on the behavior of individual sellers in the great depression, data on regional fluctuations, data on output and unemployment in Europe and Japan, as well as the data on real GNP in the U.S., studied by previous authors.

The later part of the paper examines the question of what conditions would tend to make the UEH likely to hold in individual markets. Obviously one possibility is that the UEH arises by accident; the demand function net of the response of other sellers happens to have the requisite elasticity so that the elasticity with respect to price less marginal cost is unity. But I also mention conditions under which the UEH is the logical consequence of specified assumptions. In particular, when the sellers in a market achieve the cooperative monopoly outcome and the buyers bring a predetermined amount of purchasing power to the market, then the UEH *always* holds. This conclusion remains true even when the sellers produce a variety of products and use each other's products as intermediate inputs.

### 1. The unit elasticity hypothesis

Consider the standard problem of profit maximization:

$$(1.1) \quad \max pD(p) - C(D(p))$$

where  $D(p)$  is the demand function facing the firm, net of the response of its rivals, and  $C(q)$  is the cost function. The first-order condition for profit maximization is usually expressed as

$$(1.2) \quad p + \frac{D}{D'} = C'$$

that is, marginal revenue equals marginal cost. However, an equivalent expression is

$$(1.3) \quad - (p - C') \frac{D'}{D} = 1$$

The left side of equation 1.3 contains the factor  $p - C'$ , price less marginal cost, which can be considered the part of the price captured by the fixed factors of the firm. Think of the firm as charging its customers the sum of two prices. One is a direct pass-through of marginal cost. The other is the firm's markup over marginal cost,  $p - C'$ . Then the condition says that the maximum of profit occurs when the elasticity of demand with respect to the markup part of the price is unity.

The case of a price-taking firm is a limiting case for this condition. The price taker has  $D'$  equal to infinity and achieves the condition by

setting the margin at zero, that is, by setting price equal to marginal cost. It is also worth noting how this condition works for a firm with an L-shaped cost curve, with constant marginal cost up to a certain capacity level and infinite marginal cost at capacity. In competition, the condition is satisfied either if price equals the constant element of marginal cost and output is at or below capacity, or if price is higher but output is at capacity. On the other hand, if the market is not strictly competitive, the condition permits price well above marginal cost even when output is below capacity. If marginal cost is very low because the major inputs (capital and labor) are largely fixed, then the competitive industry can only satisfy the condition by operating at capacity. Competition has a strong drive toward full resource utilization. But a firm operating in an imperfectly competitive market may well maximize profit at less than capacity output, even though marginal cost at that point is low.

The unit elasticity hypothesis holds that the elasticity of demand with respect to the markup component of the price is unity over a range of prices. That is, profit attains its maximum not just at a single price, but for all prices in some interval, perhaps fairly wide. Plainly, the firm's price and output will be indeterminate in the range where the UEH holds.

The UEH is strictly a feature of imperfect competition. In competition, the profit-maximizing price is always the market price; there is no possibility of a range of profit-maximizing prices.

A firm under the UEH will reason in the following way about a price cut: "A lower price would bring me more business. On the other hand, a lower price will generate less revenue per sale. I think the two effects offset each other, so I will not gain or lose any profit by cutting my price. I might as well stay put." Note that this logic applies *whether or not*

demand has just declined. Unless a shift in demand shifts the range where the UEH holds so that the current price is outside the range, the firm has no positive incentive to change its price.

## *2. Output is a random walk*

From the perspective of macroeconomics, the most impressive evidence in favor of the UEH (or any other theory that makes output indeterminate) is the finding that aggregate output behaves much like a random walk and not at all like a variable that tends back to a full-employment trend path. This finding is reported in Nelson and Plosser (1982), Campbell and Mankiw (1986), and Clark (1986). A similar finding for European unemployment rates is reported in Blanchard and Summers (1986).

Nelson and Plosser compare two statistical models of the log of output. One has output fluctuate above and below a deterministic trend; it corresponds to the consensus view of the equilibrium-seeking economy sketched in my introduction. The other has output evolve as a trended random walk. In the second model, a deviation in output in one year is never reversed; it has a permanent effect on later output that may be larger or smaller than the initial effect, depending on the process. Nelson and Plosser's basic tool for studying the data is the autocorrelation function of the first differences of log output. If their first model describes the data, the autocorrelations of the first difference should be

negative. After a shock displaces output from trend, future changes will be in the opposite direction as output settles back to normal. On the other hand, the second model is consistent with zero or positive serial correlation of the first differences. A pure random walk would have no serial correlation. If one favorable shock tends to follow another, then the serial correlation should be positive.

Nelson and Plosser show that the first autocorrelation of the first difference of the log of real GNP in the U.S. over the period 1909-1970 was positive. Though they find that higher-order autocorrelations are negative, they characterize their results as suggesting the first difference of log output follows a first-order moving average process with positive autocorrelation, a statistical model that is incompatible with the cycle-around-trend view and harmonious with the random-walk view. They go on to develop a model which combines both types of behavior. They demonstrate that their findings imply that the random-walk term is several times more important in explaining the year-to-year changes in real GNP.

Nelson and Plosser interpret their results as unfavorable to the consensus view of the business cycle and highly favorable to an alternative view in which output is always at its equilibrium value, and where that equilibrium is the result of the compounded effect of continuing random shocks. However, they do not investigate whether known shocks to productivity or the capital stock are large enough to explain the movements in output as a practical matter. In particular, they do not ask whether occasional large declines in output, as in the Great Depression, are the result of declines in equilibrium output.

Campbell and Mankiw (1986) provide strong confirmation of Nelson and



Plosser's findings. Campbell and Mankiw fit a general ARMA model to the first difference of the log of real GNP, using quarterly U.S. data for the postwar period. Again, the first differences tend to be positively serially correlated; Campbell and Mankiw carry out a formal test of the hypothesis that there is actually the amount of negative serial correlation implied by the cycle-around-trend model, and strongly reject that hypothesis. They show that the impact of an innovation of one percent on the prediction for the indefinite future is greater than one, instead of dying off to zero as required by the cycle-around-trend model. They note that their results are inconsistent with the current consensus about the nature of business cycles but the results are compatible with either of the type of behavior captured by equilibrium, real business cycle models or by models with multiple equilibria.

Clark's (1986) recent work also studies the first difference of the log of quarterly real GNP. His results for the autocorrelations of this variable agree with those reported by Nelson and Plosser--the lower-order autocorrelations are quite positive and the rest are close to zero. These findings are incompatible with a pure cycle-around-trend model, as Nelson and Plosser stressed. However, Clark finds that it is still possible that the cycle component accounts for around half of the quarterly innovations in the variable. He is planning to compute the longer-term implications of an innovation, to compare to Campbell and Mankiw's results. It seems likely that his random walk and random growth rate terms will dominate in the longer-run impulse reaction function, which would confirm the earlier findings. The simple cycle-around-trend model omits a dominant component of the actual behavior of real GNP.

My own work on this point, to be presented in another paper, avoids

fitting any tight parametric models to the data. Instead, it shows that there is an important low-frequency component of the behavior of the first difference of log real GNP that cannot possibly be explained by any cycle model. The reason is that the first difference of a stationary process always has power zero at frequency zero. On the other hand, the actual data are close to white noise, with only a little positive autocorrelation. The power at frequency zero is about the same as at other frequencies. For any given model of the cyclical component, it is possible to compute the maximum amount of that component that could be present in the data, and also to compute the statistical properties of the residual. No assumptions need be made about the residual at all. For realistic cycle models, the residual is the dominant element in the data. Moreover, the residual contains an important random-walk element. Application of the technique to data for Britain, France, West Germany, Italy, and Japan shows that the cycle component is even less important in those economies than in the U.S.

Blanchard and Summers (1986) examine the time-series properties of unemployment rates in the U.S. and Europe. They find serial correlation parameters for the level of unemployment of .9 and above in annual data for the postwar period in all countries save the U.S. Even in the U.S., persistence of unemployment is extreme over long stretches of history, such as 1892-1985. They conclude that a model of the unemployment rate as a random walk cannot be rejected in most cases. As they point out, the failure of the unemployment rate to return to a normal level after a perturbation is even more troubling for the standard view than is the same finding for real GNP. Shifts in technology and in productivity growth rates may have low-frequency effects on output, but should not have any

longer-term effects on unemployment, according to the standard view. Blanchard and Summers favor an explanation for the extreme persistence of unemployment based in the labor market, rather than in the product market, as in this paper. They mention a number of mechanisms in which a shock has a permanent impact. For example, a favorable demand shock may lead firms to increase the number of employees. From then on, the new employees have a voice in employment determination through their unions. Employment remains higher and unemployment lower even after the initial surge in demand has dissipated.

### *3. Other evidence of the indeterminacy of price and output*

Some of the most remarkable examples of behavior by individual firms that seem consistent only with the unit elasticity hypothesis appear in Stigler's (1947) data on product prices during the Great Depression. For two major industrial products, sulphur and nickel, nominal prices literally did not change at all from June 1929 to May 1937. From peak to trough, real GNP fell by 30 percent and nickel and sulphur sales probably fell by well over 50 percent over the same period. There was only a single important seller of nickel in the U.S. market at the time and two sellers of sulphur. Stigler's point is that theories of price rigidity based on oligopoly fail to explain the rigidity of monopoly prices. But what kind of a theory can explain the actions of a profit-maximizing

monopoly that fails to adjust its price when demand collapses? The only reasonable answer is that International Nickel did not perceive that its profit could have been increased in 1933 by setting a lower price. In other words, when the demand curve for nickel shifted inward during the Depression, International Nickel chose to move to a very different point on the new demand curve, by maintaining the same dollar price. It appears that the UEH held all along the demand curve from the point that International Nickel was on in 1929 to the very different point it was on in 1933.

A very different type of evidence pointing toward the UEH is the close connection between purely nominal changes in the economy and real output. A long tradition of empirical work, culminating in Barro's (1977) famous paper on the effect of monetary surprises, shows that a change in the money stock is usually followed by large subsequent changes in output. The explanation of that relationship has been seen as a major challenge by macroeconomists for the past two decades, but remarkably little progress has been made in developing theories grounded in good microeconomics with assumptions in accord with common sense. Theories based on contracts or other sources of rigidity have failed the first test while those based on information limitations have failed the second test.

Recent work by Poterba, Rotemberg, and Summers (1985) has documented another type of vulnerability of real quantities to purely nominal changes. They show that shifts from income taxes to commodity taxes, involving no net change in tax wedges, still have important real consequences. Wages and prices do not adjust quickly to offset the change in the tax structure.

In an economy where major sectors obey the UEH, neither of these

phenomena is a mystery. When a monetary contraction lowers output, sellers subject to the UEH do not perceive any incentive to cut prices correspondingly. Profit may have declined, but changing the price, within the range where the UEH holds, will not raise profit. If prices remain unchanged because sellers lack an incentive to cut them, then real activity must fall. In fact, the IS-LM model will do a good job of describing the implications of a monetary contraction when prices remain unchanged.

The UEH is not a theory of price rigidity, but a theory of price indeterminacy. The evidence just summarized suggests that in many instances, the price level fails to adjust to purely nominal changes, a phenomenon that is consistent with the UEH but not with models with a strong tendency toward equilibrium. Other evidence consistent with the UEH but not with other macro theories has exactly the opposite character. In some years, prices change much more than seems warranted by macroeconomic fundamentals. The most famous episode of this kind occurred in the U.S. and Britain in the aftermath of World War I. Figure 1 shows the tremendous swing in prices in the U.S. from 1919 through 1922. It also shows the relatively modest movements in output. The price level rose sharply in 1920 and then fell precipitously to below its 1919 level in 1921. In Britain, the rise and fall of the price level was even more pronounced. These price movements have generally been ignored by the proponents of models of price rigidity, who tend to begin their histories a few years later during the Great Depression. The contrast between the quick decline of prices in the contraction of 1920-21 and the slow deflation that accompanied the prolonged and deep contraction of 1928-33 presents major problems for any deterministic theory of price and output determination.

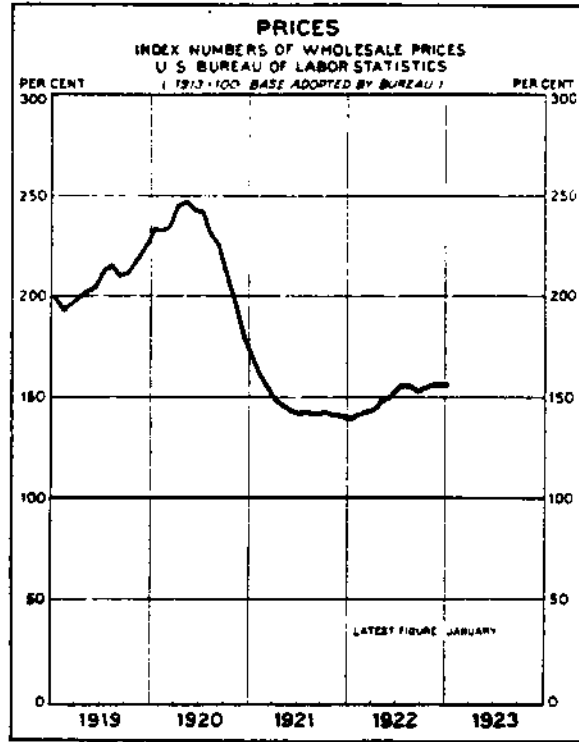
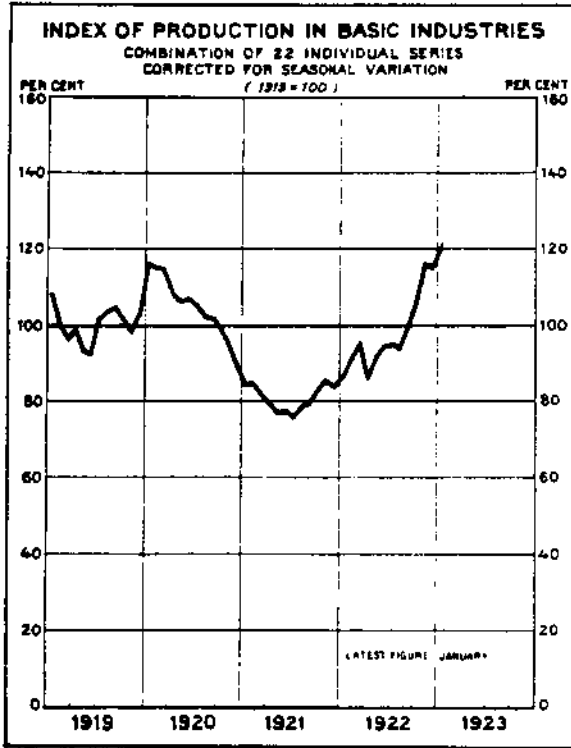


Figure 1. Output and prices in the U.S.,  
 1919-1922. Source: Federal Reserve  
Bulletin, 9:286, March 1923

#### 4. Analytical models that imply the unit elasticity hypothesis

All theories of profit-maximizing behavior on the part of firms imply that they perceive unit elasticity of demand with respect to the markup part of the price. As I showed in section 1, this proposition is nothing more than a restatement of the first-order condition for profit maximization. One cannot dismiss the UEH on the grounds that it understates the numerical value of the elasticity of demand. When the UEH fails, it is because the elasticity of demand is rapidly increasing at the point of profit maximization, not because the elasticity exceeds one.

Most of what I want to say about the theory behind the UEH relates to the case of a technology with constant marginal cost and a capacity constraint. Suppose that the unrestricted or long-run cost function is

$$(4.1) \quad (\lambda + r) q$$

where  $\lambda$  is the marginal labor and materials cost of a unit of output,  $r$  is the annual carrying cost of a unit of capacity, and  $q$  is the amount of output. Suppose further that the restricted or short-run cost function is

$$(4.2) \quad \begin{array}{ll} \lambda q & \text{for } q \leq k \\ \infty & \text{for } q > k \end{array}$$

Marginal cost is  $\lambda$  for output below capacity,  $k$ , and can be taken to be any number greater than  $\lambda$  when output is at  $k$ .

Let the demand function facing the industry be  $D(p,x)$ , where  $x$  is a random variable whose value is not known at the time that capacity is chosen. I will also make use of the Marshallian form of the demand function,  $p = \phi(q,x)$ , defined by  $D(\phi(q,x),x) = q$ .

### *Competitive equilibrium*

Once capacity,  $k$ , has been chosen and  $x$  is drawn, the price is

$$(4.3) \quad p = \psi(k,x) = \max(\phi(k,x), \lambda)$$

The price is either equal to short-run marginal cost,  $\lambda$ , and output is less than capacity, or price is the level necessary to keep the quantity demanded down to the level of capacity. At capacity planning time, a competitive industry chooses enough capacity to set expected profit to zero. Expected profit is

$$(4.4) \quad E\{(\psi(k,x) - \lambda)k\} - rk$$

Here I have used  $k$  in place of  $q$  because whenever  $q$  is less than  $k$ ,  $\psi - \lambda = 0$ , so it does not matter what it multiplies. Setting expected profit equal to zero and dividing by  $k$  gives the condition for zero expected profit in terms of price behavior:

$$(4.5) \quad E(\psi(k,x)) = \lambda + r$$

that is, expected price equals long-run unit cost. Note that the highly



nonlinear relation between  $x$  and profit does not call for any special treatment in this zero-profit condition. It is clear from equation 4.5 that a competitive industry will plan to be at full capacity utilization a reasonably large fraction of the time. Whenever output is less than capacity, the industry sustains a loss of  $rk$ . If demand is fairly price-elastic and the distribution of  $x$  puts little weight on high values of demand, then the industry will have to be at capacity quite frequently in order to earn enough revenue to cover the cost of its capacity,  $rk$ . Only if there is some chance of a very high value of  $\psi(k,x)$  can the competitive industry expect to cover its capacity costs without operating at capacity much of the time.

The unit elasticity hypothesis is effectively precluded by the assumption of competition. When the market price exceeds  $\lambda$ , all firms face a strong incentive to produce at capacity. Their profit functions are not even locally flat with respect to their output. When the market price is exactly  $\lambda$ , the output of each firm is indeterminate, but price is not.

### *Monopoly*

The monopolist faced with the same demand and cost conditions will choose the price that maximizes profit once demand is known, and will choose the capacity that maximizes expected profit. When the monopolist faces sufficiently strong demand so that his output reaches capacity, his revenue is the same as the revenue of a competitive industry

with the same capacity. However, when output is below capacity, the monopolist earns much higher revenue, because the monopoly price exceeds the competitive price. A monopoly with the same capacity as a competitive industry will have lower capacity utilization, because its high price will result in lower sales. But a monopoly will also choose a lower level of capacity than will a competitive industry, which will raise its capacity utilization rate. As a general matter it does not seem possible to determine whether a monopoly has a lower or higher capacity utilization rate than does a competitive industry with the same demand and cost conditions. The monopolist does not have an incentive to plan chronic excess capacity. Like the competitive industry, the monopolist cannot be governed by the unit elasticity hypothesis at times when output is at capacity. Profit is an increasing function of output at such times, and there is no possibility of indeterminacy except in the borderline case. However, the UEH may come into play in those times when the monopoly is selling less output than its capacity. At those times, the elasticity of demand with respect to the markup part of the price must be unity. It is quite possible that this condition will hold over a range of prices and outputs. I shall have more to say on this point shortly.

### *Shared monopoly with free entry*

A market structure that seems to fit the facts better than do competition or monopoly in a number of industries is the following: All the firms in an industry coordinate their activities in such a way as to achieve the monopoly outcome, given their capacity. However, they are

unable to limit entry. A potential entrant correctly anticipates that he will join the shared monopoly once he has built his plant. Hence entry occurs up to the point of zero expected profit.

Shared monopoly is one of the many outcomes consistent with repeated-play games of oligopoly. However, a much stronger and more persuasive model of shared monopoly relies on Anderson's (1986) theory of quick-response equilibrium. In the market governed by Anderson's theory, rivals learn instantly about each other's price changes. They can respond so quickly that a price-cutter does not earn any significant extra profit by undercutting the other sellers. Rather, the rivals match any price cuts that occur. Customers see a simultaneous reduction of all prices. Similarly, each seller has to decide whether to match a price increase. If all sellers match an increase, it goes into effect; if not, those who did match the increase rescind their increases with no lost profit. Price changes are made, in effect, by a voting system. It takes one vote to cut the price and a unanimous vote to increase the price. The equilibrium is easy to see: It sets a price which maximizes the profit of the lowest-cost seller, subject to all other sellers charging the same price. At that price, there are no votes for a price cut and  $N-1$  votes for a price increase. At any higher price, there is at least one vote for a cut; at any lower price, the vote is unanimous for an increase.

When all sellers have the same cost, Anderson's quick-response equilibrium is simply monopoly, with output allocated by consumer behavior in the face of equal prices from all sellers. As he points out, this conclusion is profoundly unsettling for believers in competitive markets and their attendant efficiencies. The act most emblematic of competition—one seller instantly cutting his price to match a competitor's

--is the undoing of competition. Consistent price-matching supports the monopoly outcome.

The sellers in a quick-response equilibrium do not have any way to deter entry, even though they have a foolproof way to control the entrant's price. Entry will occur until all sellers have zero expected pure profit. The level of capacity at that point will exceed the capacity of either a competitive market or a monopoly. Capacity utilization will be below the level of competition or monopoly. It is even possible that capacity will be built in excess of the highest possible level of output. These conclusions are spelled out in a series of propositions:

*Proposition 1.* The shared monopoly, free entry (SMFE) equilibrium has capacity in excess of the competitive level and capacity utilization below the competitive level.

Proof: Figure 2 shows the profit per unit of capacity in competition and monopoly. For competition,  $\pi_C/k$  declines monotonically—added capacity reduces the revenue obtained in all circumstances where capacity is fully utilized and does not add to revenue when there is excess capacity. The competitive level of capacity,  $k_C$ , is shown at the point where  $\pi_C/k$  crosses the horizontal axis. Monopoly profit per unit of capacity,  $\pi_M/k$ , declines monotonically for the same reason.. The SMFE equilibrium capacity,  $k_S$ , occurs where  $\pi_M/k$  crosses the horizontal axis. Finally, because monopoly output never exceeds the competitive output, capacity utilization in the SMFE equilibrium must always be below the competitive level.

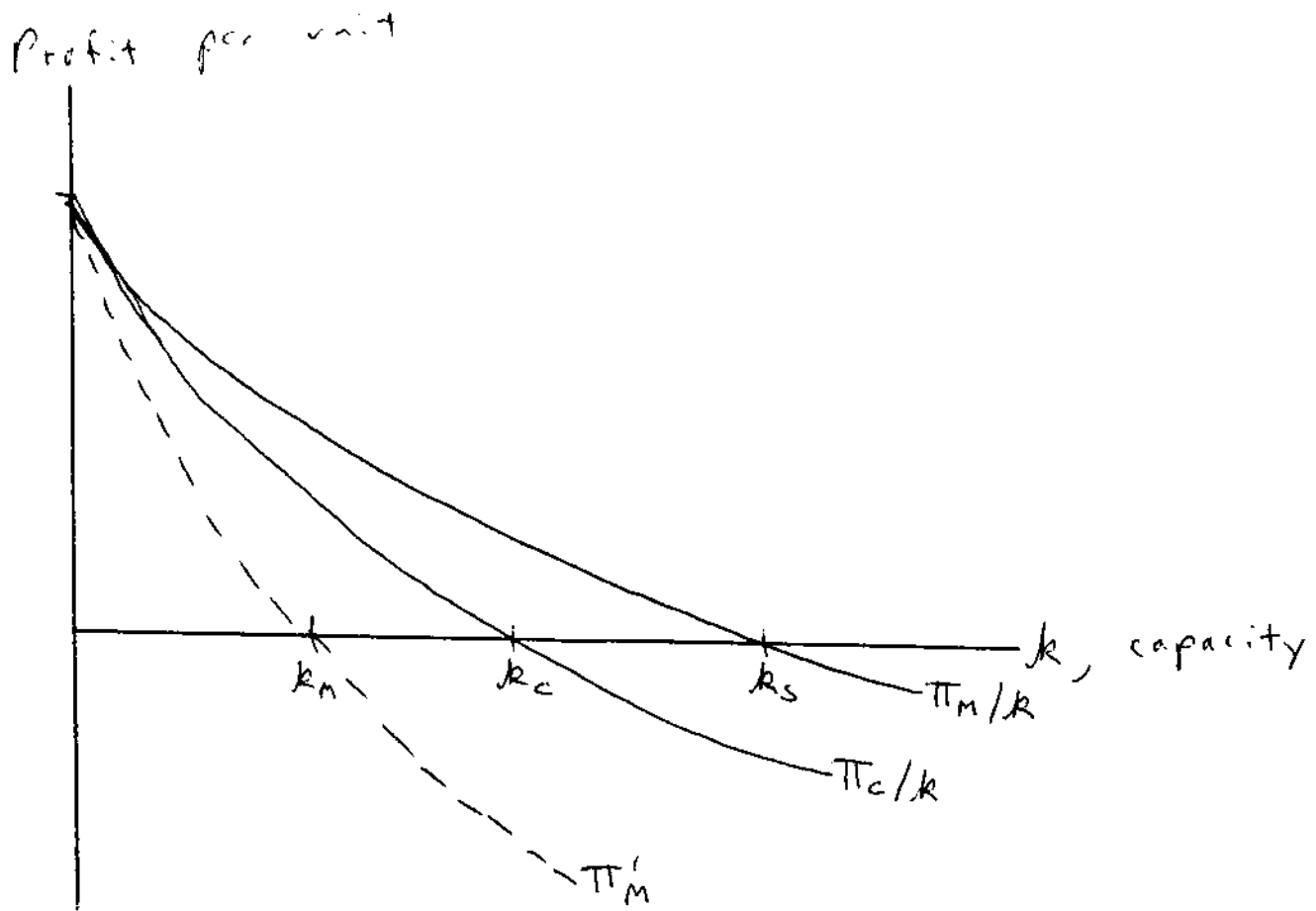


Figure 2. Profit per unit of capacity

*Proposition 2.* The SMFE equilibrium has capacity in excess of the monopoly level and capacity utilization below the monopoly level.

Proof: The dashed line in Figure 2 shows marginal profit per unit of capacity for the monopolist; the monopoly level of capacity,  $k_M$ , is shown at the point where that line crosses the horizontal axis. Plainly  $k_M < k_S$ . Monopoly and shared monopoly have the same level of output when  $q < k_M$ , and shared monopoly has less than full capacity utilization in some circumstances when monopoly is at full capacity, so the SMFE equilibrium must have a lower capacity utilization rate.

*Proposition 3.* It is possible for the SMFE equilibrium to have substantial redundant capacity; that is, for capacity to exceed the highest possible level of output by a considerable margin.

Proof: The proof is by example. Suppose that there is no uncertainty about demand and  $D(p) = 1-p$ . Short-run marginal cost,  $\lambda$ , is .2 and the marginal cost of capacity,  $r$ , is .1. The monopolist with unlimited capacity would set a price of .6 and sell .4 units. Revenue would be .24 and revenue net of short-run marginal cost would be .16. The level of capacity needed to achieve zero profit would be  $.16/.1$ , or 1.6 units of capacity. Only .4 of these units would be used, so capacity utilization would be 25 percent.

Chronic excess capacity is the normal state of affairs in an SMFE equilibrium. Full capacity utilization is seldom if ever achieved. Most of

the time, the industry is at an interior point where the elasticity of demand with respect to  $p-\lambda$  must be unity. Further, it does not strain credulity that they may face an interval over which that elasticity is unity, so that the industry price and output is indeterminate. It is easy to solve for the demand function implied by the UEH. It is

$$(4.6) \quad D(p) = \frac{D}{p-\lambda}$$

The elasticity of demand with respect to  $p$  is

$$(4.7) \quad \epsilon = \frac{p}{p-\lambda}$$

The elasticity is a declining function of price; it exceeds unity and approaches infinity as  $p$  approaches  $\lambda$ . I would argue that an elasticity of demand mildly decreasing in price is completely plausible. In particular, if the demand for a product is the sum of the demands of various agents, each of which has its own, constant price elasticity, then the price elasticity of the sum will be decreasing. As the price rises, the composition of demand will shift toward the less elastic agents, and the aggregate elasticity, which is the weighted average of the individual elasticities, will rise.

It should be apparent that the unit elasticity hypothesis cannot hold over a range of prices that is close to short-run marginal cost,  $\lambda$ . Demand would have to be highly elastic for prices close to  $\lambda$  and much less elastic for prices well above  $\lambda$ , in order for profit to be equal at the two prices. Rather, the UEH is only plausible for a range of prices that starts well above marginal cost; that is, in cases that are far from competitive.

Figure 3 illustrates the way that the elasticity of demand would have to decrease with price in order to satisfy the UEH. Each line is labeled with the elasticity of demand,  $\epsilon$ , at the price  $p = 1$  and also with the markup ratio,  $\mu$ , implied by that elasticity (the markup ratio is  $\mu = \epsilon/(\epsilon-1)$ ). The top line shows that the elasticity of demand would have to vary from 2 to 6 in order to make the price indeterminate over the range from 1.3 to 0.8. In the middle of that range, at  $p = 1$ , the elasticity would be 3 and price is 1.5 times marginal cost. In other words, in an industry where price was observed to be around 1.5 times marginal cost, the elasticity of demand would have to be quite sharply decreasing with price in order for the UEH to hold. On the other hand, the lowest line shows that the elasticity could be very nearly constant if the firm is far from competitive. If the elasticity is 1.2, so that price is marked up 6 times over marginal cost, then the UEH seems eminently plausible over quite a wide range of prices.

I draw the following conclusion: The UEH cannot possibly hold in a competitive industry; instead, the price set by each firm is made determinate by strong forces. In a non-competitive industry where each firm perceives its price elasticity of demand to be well above one, so that its price is marked up over marginal cost by less than a factor of two, the UEH is possible but unlikely. In such an industry, the elasticity of demand would have to decline quite rapidly with price in order to make price indeterminate. But in a non-competitive industry with a much lower elasticity and a high markup ratio, the UEH seems likely to hold as a reasonable approximation. Thus, industries found to have high markup ratios are the candidates for the UEH to hold, with the consequent indeterminacy of output and price.



# Elasticity of demand

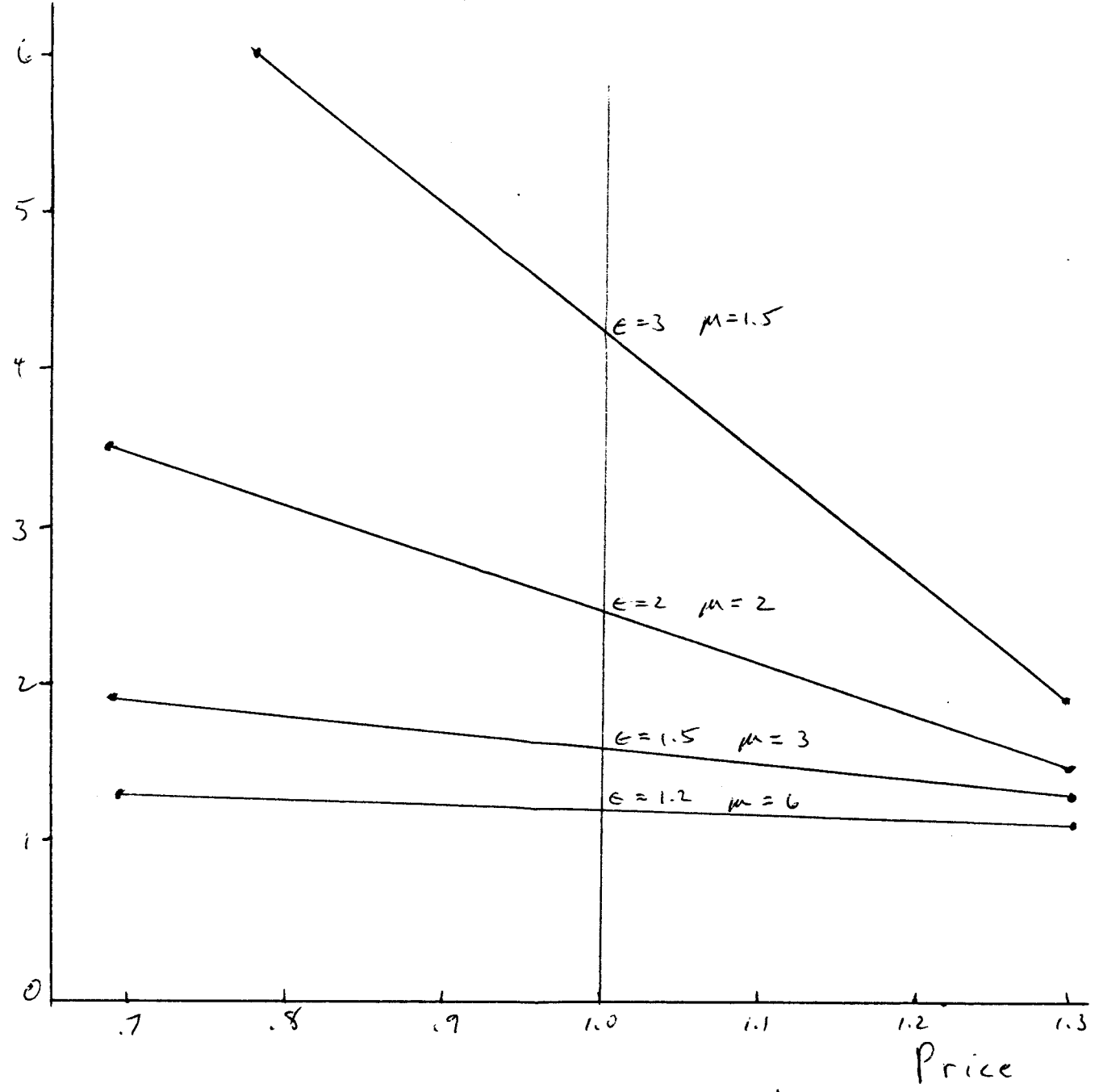


Figure 3. Elasticities of demand consistent with the unit elasticity hypothesis

Figure 4 examines a closely related question. Suppose the elasticity of demand is constant, instead of declining with price as required by the UEH. How much profit does the firm lose as a result of setting a price different from the unique profit-maximizing one? In the strongly non-competitive case with the markup ratio of 6, profit is essentially perfectly flat. Setting price 30 percent below the optimum of 1 costs 1.8 percent of profit; setting it 30 percent too high costs 0.7 percent of profit. The consequences of price-setting errors for more competitive firms are much more severe, as shown by the lower curves in Figure 4.

#### *Theoretical models where the UEH always holds*

The simplest model where the UEH is inevitable is the following: There is a single firm, endowed with the ability to produce unlimited quantities of good A at zero marginal cost. There are many consumers, each endowed with a fixed amount of good B. The owner of the firm is interested in consuming B, while consumers are interested only in consuming A. Because of their number, consumers are price-takers. The monopolist posts a price,  $p$ , representing the number of units of B he requires to give up one unit of A. The monopolist consequently faces a demand function,

$$(4.8) \quad D(p) = B/p$$

where  $B$  is the aggregate endowment of good B. His objective is to maximize his revenue, that is, the number of units of B that he acquires in trade. Revenue is  $pD(p) = B$ , the same for all prices. The demand

Lost profit, percent

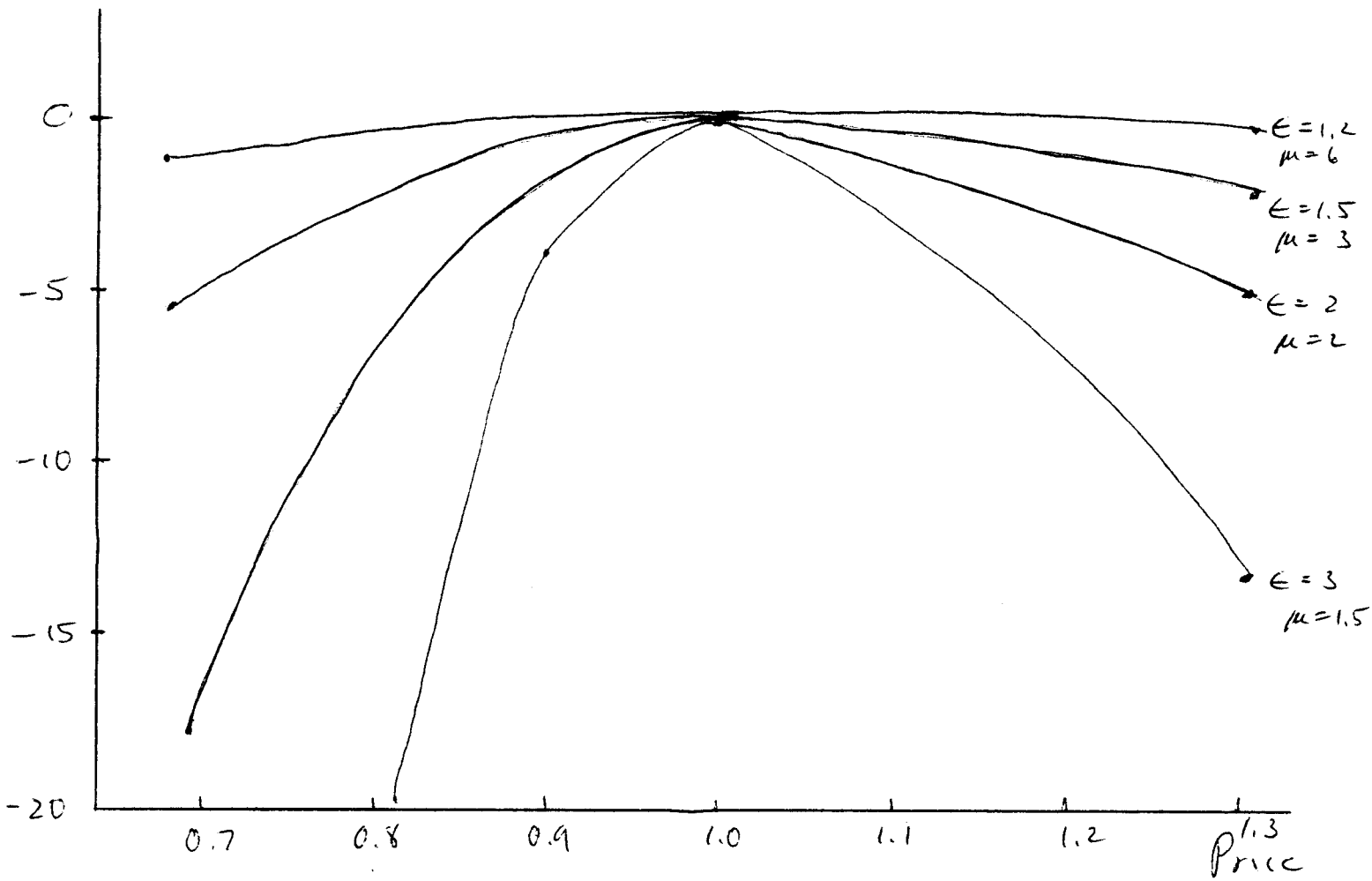


Figure 4. Profit as a function of price under various demand elasticities.

schedule is unit-elastic everywhere and the UEH holds for all prices. The output of the monopolist is  $B/p$ . Although the monopolist is indifferent between, on the one hand, low output and high price, and, on the other hand, high output and low price, consumers are far from indifferent. Consumers strongly prefer low price and high output. The efficient outcome calls for enough output of good A to saturate consumers, but there is no guarantee this will happen.

Readers of Hart's (1982) seminal paper will recognize this setup as a special case of the situation in one of his local markets. It is special in three ways. First, there is only one seller. Under Hart's assumption of Cournot equilibrium with several sellers, the indeterminacy could not arise. Second, consumers do not want to consume the good they are endowed with. If they did, as in Hart's model, their demand for the produced good would not be everywhere unit-elastic. Third, there is no labor in this special version. Hart's point was to show that imperfect markets generally had an equilibrium with less than full resource utilization. Under his assumptions, the equilibrium is determinate. Under my special assumptions, the possibility of under-utilization remains, but the equilibrium is indeterminate.

Cooper and John (1985) have studied a related model. Consumers do not have a physical endowment of any good, but derive abstract purchasing power from employment in another market. There are Cournot sellers in the local market where the consumer dwells, and these sellers are willing to trade their output for the consumer's purchasing power. Again, the special case of monopoly brings about the type of indeterminacy associated with the UEH--the seller can capture all of the purchasing power in his market no matter what price he sets, so he is indifferent.

However, Cooper and John study a very different type of indeterminacy, one unrelated to the structure of the product market. Because the consumers in one market derive their purchasing power by working in another market, and the consumers in the second market derive their purchasing power from working in the first, an equilibrium can occur at any level of output, as long as the two markets have the same output. Cooper and John's indeterminacy is the same as the one that arises in the Keynesian cross model in the special case where the marginal propensity to consume is unity and there is no autonomous spending. It is a fragile indeterminacy in that it disappears if consumers have any way to finance consumption other than through current work. Access to borrowing and lending markets makes the marginal propensity to consume quite a bit less than one and rules out Cooper and John's indeterminacy. The UEH will arise in all cases where the consumers in a market are committed to spend a given amount of purchasing power in the market, and where all the sellers cooperate in extracting the purchasing power. This conclusion holds even if there is a multiplicity of products and if the producers use each other's products as inputs. Ultimately, the profit of the producers is exactly the purchasing power of the consumers, and all of this purchasing power can be extracted with any set of prices.

## 5. Empirical evidence on market imperfections

The last section showed that indeterminacy of price and output is most likely to arise when firms perceive a distinctly downward-sloping demand schedule. Although it is true that every profit-maximizing firm, independent of market structure and demand elasticity, operates at a point where the elasticity of demand with respect to the markup part of the price is unity, it is only plausible for it to be near unity for a range of prices if the markup part is substantial, that is, if competition is far from perfect.

My own research (Hall, 1986) studied the markup of price over marginal cost, using a direct measure of marginal cost. Marginal cost is inferred from the actual change in labor cost that occurs from year to year. The ratio of price to marginal cost is expressed as  $1/(1-\beta)$ , so  $\beta=0$  implies zero markup (perfect competition) and positive values of  $\beta$  imply positive markups. Because data restrictions make it necessary to measure the output of an industry as its real value added and its price as the corresponding deflator, the interpretation of  $1/(1-\beta)$  is the markup of the deflator over marginal labor cost, not the markup of the product price over total marginal cost. I show that the latter is  $1/(1-(1-\mu)\beta)$ , where  $\mu$  is the factor share of materials and other purchased inputs.

Table 1 shows the estimated values of  $\beta$  from my study, in the first column, and the estimated markup ratios of product prices over full marginal cost, in the second column. None of the ratios is as high as 2.0, and few exceed 1.5. On this basis, there is relatively little support for the UEH. However, these estimates make the strong assumption that firms are price-takers in their input markets. A very different view

Table 1. Estimates of markups by industry

SIC code	Description	Est. $\hat{\beta}$	Industry markup $\frac{1}{1-(1-\mu)\hat{\beta}}$	Full markup $\frac{1}{1-\hat{\beta}}$
20	Food & beverages	0.837	1.33	6.13
21	Tobacco	0.635	1.39	2.74
22	Textiles	0.556	1.32	2.25
23	Apparel	0.275	1.15	1.38
24	Lumber	0.591	1.48	2.44
25	Furniture	0.271	1.13	1.37
26	Paper	0.661	1.47	2.95
27	Printing and pub.	0.584	1.43	2.40
28	Chemicals	0.831	1.62	5.92
29	Pet. refining	0.874	1.10	7.94
30	Rubber	0.289	1.14	1.41
31	Leather	0.405	1.25	1.68
32	Stone, clay, glass	0.536	1.38	2.16
33	Primary metals	0.512	1.29	2.05
34	Fab. metals	0.291	1.15	1.41
35	Non-elec. mach.	0.273	1.16	1.38
36	Elec. mach.	0.301	1.18	1.43
38	Instruments	0.251	1.16	1.34
39	Miscellaneous	0.485	1.28	1.94
48	Communication	0.529	1.68	2.12
49	Elec. gas, sanitary	0.965	1.94	28.57

Source: Hall (1986), Table 2.

emerges under a different and more realistic model of the relation between upstream and downstream producers. In each of the industries in Table 1, many of the producers are vertically integrated with upstream producers, and many others have explicit or implicit contracts. In all of those cases, a reasonable model is that the upstream and downstream producers cooperate with each other. In that case, the appropriate concept of  $\beta$  is for the combined operation of the upstream and downstream firms. That is,  $\beta$  can be measured by comparing the price set by the downstream firm to the marginal cost incurred by the two firms, not counting the output of the upstream firm. It is easy to show that the markup parameter,  $\beta$ , for the combined operation, measuring the markup of actual product price over all components of marginal cost, is

$$(5.1) \quad \beta = \hat{\beta}_D + \mu_D(\beta_U - \hat{\beta}_D)$$

where  $\hat{\beta}_D$  is the markup parameter for the downstream firm, as estimated in my work (markup of the deflator over marginal labor cost),  $\mu_D$  is the share of materials in the downstream industry, and  $\beta_U$  is the full markup parameter for the upstream firm. If  $\hat{\beta}_D$  is close to  $\beta_U$ , or  $\mu_D$  is not too high, then  $\hat{\beta}_D$  will be a good estimate of  $\beta$ . The third column of Table 1 shows markups calculated as  $1/(1-\hat{\beta})$  for each industry. On this basis, many of the industries are easily in the range where the UEH is plausible.



## 6. *Concluding remarks*

In competitive or nearly competitive markets, there is a strong force toward full utilization of capacity and full employment of labor. Each seller ignores the influence of his own choice of output on the market price. As long as the price is above short-run marginal cost, there are substantial profits to be made by expanding output to its limit. Excess capacity will be found only if price falls all the way to short-run marginal cost. In an economy dominated by competitive markets, total output will always be close to an equilibrium level determined by the supply side of the economy. Such an equilibrium level will grow smoothly over time. Temporary departures from the level will be reversed speedily by competitive forces.

In an economy with distinctly non-competitive markets, the forces pushing output to its capacity level are weaker or absent. Each seller, or group of sellers effectively operating in concert, will be cautious about expanding output because the increased volume will bring lower prices. In one important case, the two effects exactly offset each other, and profit is the same over a range of levels of output and corresponding prices. In that case, the unit elasticity hypothesis holds. The UEH is bound to hold whenever cooperating sellers as a group are guaranteed to capture all of the purchasing power brought to a market by the customers in the market. In that case, the collective profit of the sellers must always equal the purchasing power available in the market, quite independent of the prices set for the various products sold in the market. More generally, the UEH will arise as a reasonable approximation in any market where sellers have enough market power so that their prices are marked up by double

marginal cost or more. Profit maximization requires that the elasticity of demand with respect to the markup part of the price be equal to unity. When the markup part is large (at least as large as marginal cost), then it does not vary too rapidly over a range of prices around the maximum of profit. Consequently, the elasticity must be near unity over that range, and so profit must be close to its maximum over the range. Hence, price is indeterminate or nearly indeterminate.

Empirical evidence on the importance of the UEH in the U.S. and other economies falls into two major categories. First, the time series behavior of real GNP and related variables is inconsistent with the competitive model (assuming that the determinants of supply grow along smooth trends) and consistent with the UEH or any other model where output is indeterminate. Real GNP does not return to normal after the typical shock. Its time series properties are dominated by a random walk element, which is inconsistent with a return to normal after a shock. The other evidence supporting the UEH has emerged from studies that inferred the slope of the demand schedule facing the firm either directly or by looking at the ratio of price to marginal cost. Both approaches agree that an important fraction of industries are far from competitive, and are within the numerical range where the UEH is plausible.

## *References*

Robert Anderson, "Quick-Response Equilibrium," Departments of Economics and Mathematics, University of California, Berkeley, February 1985.

Robert Barro, "Unanticipated Money Growth and Unemployment in the United States," *American Economic Review* 67:101-115, March 1977

Olivier Blanchard and Lawrence Summers, "Hysteresis and the European Unemployment Problem," MIT, Harvard, and the National Bureau of Economic Research, February 1986

Timothy Bresnahan, forthcoming chapter in the Handbook of Industrial Organization

John Campbell and Gregory Mankiw, "Are Output Fluctuations Transitory?" Princeton, Harvard, and the National Bureau of Economic Research, January 1986

Peter Clark, "The Shape of U.S. Business Cycles," Graduate School of Business, Stanford, January 1986

Robert E. Hall, "The Relation between Price and Marginal Cost in U.S. Industry," Working Paper 1785, National Bureau of Economic Research, January 1986

Oliver Hart, "A Model of Imperfect Competition with Keynesian Features," *Quarterly Journal of Economics* 97:109-138, February 1982.

Charles Nelson and Charles Plosser, "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications," *Journal of Monetary Economics* 10:139-162, September 1982

James Poterba, Julio Rotemberg, and Lawrence Summers, "A Tax-Based Test for Nominal Rigidities," National Bureau of Economic Research Working Paper 1627, June 1985

George Stigler, "The Kinky Oligopoly Demand Curve and Rigid Prices," *Journal of Political Economy* 55, October 1947, reprinted in George Stigler, *The Organization of Industry* (Homewood, Illinois: Richard D. Irwin), 1968, pp. 208-234.