The Genealogy of the Labor Hoarding Concept

Jeff E. Biddle
Dept. Of Economics
Michigan State University

May 2013

Abstract

The modern concept of labor hoarding emerged in early 1960s, and soon became a standard part of mainstream economists’ explanation of the working of labor markets. The concept represents the convergence of three important elements: an empirical finding that labor productivity is procyclical; a framing of this finding as a “puzzle” or anomaly for the basic neoclassical theory of the firm, and a proposed resolution of the puzzle based on optimizing behavior of the firm in the presence of fixed costs of hiring, firing, and training workers. The paper recounts the history of each of these elements, and how they were woven together into the labor hoarding concept. Each history involves people associated with various research traditions, and motivated by an array of questions, many of which were unrelated to the questions that the modern idea of labor hoarding was ultimately created to address.
By the end of the 20th century, the concept of “labor hoarding” had become an accepted part of mainstream economists’ explanations of the workings of labor markets and of the relationship between labor productivity and economic fluctuations. For example, the 1999 edition of a popular undergraduate labor economics text included labor hoarding in its glossary of key concepts, defining it thusly: “The practice by which firms retain more workers during recessions than would be technically necessary, specifically overhead workers such as executives, managers and skilled workers on whom the firms have spent large sums to recruit and train (McConnell, Brue, and MacPherson 1999). The phrase “labor hoarding” had entered this text in its 1989 edition, while the 1986 edition contained essentially the same explication of the phenomenon without the use of the phrase. In all editions, labor hoarding was discussed in a section dealing with cyclical movements in aggregate labor productivity. In the 1980s editions of the undergraduate labor economics texts by Flanagan, Smith and Ehrenberg and by Hamermesh and Rees, labor hoarding was discussed in a chapter on the demand for labor, although in both these texts the link between the labor hoarding phenomenon and procyclical movements of labor productivity was also highlighted. All three texts noted that labor hoarding was a profit maximizing behavior on the part of employers, caused by the existence of fixed costs of hiring, firing, and training workers. In macroeconomics, the 1981 edition of widely used intermediate-level textbook by Dornbusch and Fisher italicized the phrase labor hoarding in a section dealing with “Output and Productivity.” The section began with the observation that while the neoclassical analysis of earlier chapters implied that labor productivity should move countercyclically, measured aggregate productivity actually moves procyclically. Thus an “essential modification of the neoclassical analysis is required,” and a recognition of an incentive to hoard
labor in response to costs of adjusting the labor force was a crucial element of this modification (Dornbusch and Fisher 1981, p. 369).

What I call in this paper the modern concept of labor hoarding was first fully articulated in the early 1960s, and involves the conjunction of three key elements:

A fact: Measured labor productivity (output per worker or per manhour) is procyclical, rising during expansions and falling during contractions. This fact was also presented in an alternative form as a short run elasticity of employment with respect to output of less than unity; i.e., the amplitude of employment movements over the cycle is smaller than the amplitude of output movements over the cycle.

A perceived contradiction between that fact and a widely held theory: In the basic neoclassical theory of the firm in a competitive industry, short run fluctuations in demand are met by increases or decreases in the labor input, with the capital stock unchanged. Given the assumption of diminishing marginal returns to labor, this would lead average labor productivity to move countercyclically, decreasing when output increased, and increasing when output decreased.

A possible explanation, rooted in optimizing behavior on the part of firms, and thus consistent with the larger neoclassical paradigm: Costs of hiring, firing, and training employees make it optimal for employers maximizing profits over a longer-term time horizon to keep workers employed in the face of short-run drops in demand. This dampens the amplitude of employment fluctuations in response to demand fluctuations, leading average output per worker (measured productivity) to fall when demand falls.
In a 1985 article that made an influential contribution to the empirical literature on labor hoarding, Jon Fay and James Medoff offered a brief history of the emergence of the labor hoarding concept. Readers were told that until 1960, economists “generally adopted the assumption that the technical production function exhibits decreasing average product of labor in the short run, and that labor is freely variable.” This led to the implication that “labor productivity should rise during downturns and fall during upturns.” Thor Hultgren “tested this belief” empirically in 1960, and found contradictory evidence. In 1964, Robert Solow emphasized that this was a “puzzle” that it was very important for economists to solve, and a substantial literature on the puzzle subsequently developed. According to Fay and Medoff, economists reacted to this either by modifying their prior theoretical assumption, or by questioning the reliability of the empirical finding. On the theoretical side, some economists abandoned their prior beliefs regarding the nature of the short run production function, but the more popular approach was to reject the assumption that labor is a freely variable factor, and embrace the labor hoarding hypothesis.

Fay and Medoff’s account correctly identifies key individuals and events, but the larger picture it conveys is quite inaccurate. It is not my purpose, however, to critique Fay and Medoff’s review of the literature as they found it in 1985. They were not writing as historians, and providing a reliable narrative of the emergence of the labor hoarding concept was decidedly not the goal of their research. I only mention their account because I believe that it portrays the events surrounding the emergence of the modern concept of labor hoarding in the way that most practitioners would expect that they had occurred: an empirical study was conducted to test a well known and widely held theoretical hypothesis; the hypothesis appeared to fail the test, and the research community embarked on efforts, both theoretical and empirical, to resolve the
apparent anomaly. I also believe that this rather orderly account of the appearance of and reaction to an anomalous empirical finding in economics stands in interesting contrast to the more complex story I present here. As I will describe below, each of the three elements that were woven together into the modern labor hoarding concept in the early 1960s has a history of its own. And it turns out that the various people involved these three histories were motivated by a diverse set of purposes, worked in a variety of theoretical and methodological traditions, and were often not mainly, if at all, concerned with the questions which the labor hoarding concept was ultimately used to address.

The Discovery of the Fact of Procyclical Labor Productivity

For the first fifty years of the 20th century, the conventional wisdom among economists was that labor productivity was countercyclical. There were three main arguments offered in support of this belief, none of which involved the neoclassical assumption of diminishing returns to increases in a homogenous labor input applied to a fixed capital stock. The first argument was that the rising demand for labor occasioned by an economic expansion would force employers to hire lower quality workers, while with the onset of recession the least productive workers would be the first to be discharged. The second was that the overtime hours typically demanded of workers during economic expansions would push those workers past their point of maximum efficiency. The third held that workers themselves would be motivated to work harder when they feared the prospect of job loss, as in a recession, and would slack off when labor markets were tight, employers loathe to let workers go, and good alternative jobs readily available.

Wesley Mitchell, in his influential 1913 book *Business Cycles*, phrased the arguments this way:
Where humanitarian motives are not allowed to interfere with business policy, the less efficient employes (sic) are the first to be discharged after a crisis. Hence the relatively small working forces of depression are the picked troops of the industrial army. When a revival has grown into full prosperity, on the contrary, employers are constrained to accept any help to be had. They must take on men who are too old, and boys who are too young, men of irregular habits, men prone to malinger, even the chronic trouble makers . . . A deterioration of the average efficiency of the working force inevitably follows.

. . . Now overtime labor is especially expensive to employers, not only because it often commands extra rates of wages, but because it is tired labor . . . (A)fter a time all hours of every day find the men less alert and less energetic – unable to accomplish as much work per hour as in less busy seasons. Moreover, the quality of output declines as nerves become fatigued . . .

Quite apart from this difficulty of overtime, men cannot be induced to work at so fast a pace when employment is abundant as when it is scarce. Employers complain that in good times their men slow down; employes complain that in dull times they are speeded up. (Mitchell 1913, pp. 477-478)

Mitchell’s interest in the behavior of labor productivity over the business cycle arose from his desire to construct a theory, or as he phrased it, an “analytical description” of the cycle, in which the circumstances which caused economic crisis and recession were a necessary consequence of the profit seeking actions of firms in response to economic expansion, while
economic expansions grew out of the circumstances created by the reactions of businessmen to recession. The alleged upward pressure on labor cost per unit of output that took place during economic expansions as labor efficiency declined, and the decline in cost per unit during recession as labor efficiency increased, were important mechanisms in the theory.

Mitchell’s methodological convictions compelled him to find empirical backing, if possible, for each element of his analytical description. As of 1913, however, the only broad based empirical evidence he could offer for his speculations regarding countercyclical labor productivity spoke to the importance of the third argument:

The most trustworthy body of evidence on the subject, however, is contained in the special report made by the Commissioner of Labor in 1904 on *The Regulation and Restriction of Output*. The evidence is the more convincing because the influence of business conditions upon the efficiency of labor was not a subject of inquiry. Nevertheless, manufacturers and foremen, trade-union officials and manual workers both within and without the ranks of organized labor called attention time after time to the fact that the pace of work was slower in the flush times of 1900-02 than it had been in the dull times of 1894-96. In different phrases they all gave the same explanation – men are less afraid of discharge when business is good. (Mitchell 1913, p.478)

Mitchell went on to quote one employer’s testimony that the impact of the changing worker attitudes had a much greater negative impact on productivity during a boom than did the lowering of hiring standards.
The arguments employed by Mitchell in support of the assumption of countercyclical labor productivity remained common currency among economists for the next few decades. For example, in a 1922 article on “Personnel Problem and the Business Cycle,” Paul Douglas explained how the increased demand for labor occasioned by prosperity was met by hiring less efficient workers. Douglas also spoke of the “breakdown of the fear incentive” during vigorous expansions, which led to more absenteeism, greater turnover, and “served to make the worker less efficient on the job itself.” As a result, output per man-hour fell. (Douglas 1922, p. 18)

During slow economic times, the men who remain employed were fairly skilled operators, and “the fear of unemployment serves . . .to make men work harder than they did before the period of depression.” (Douglas 1922, p. 21) What makes Douglas’s discussion noteworthy for our purposes is that he supported his arguments with actual productivity statistics reported by various employers during the boom years at the end of World War I and the short, severe recession of 1921. Data from building trade employers in Chicago showed the number of bricks laid per 8 hour day falling from 1500-1600 in 1914-16 to 700-800 in 1918-20, and statistics from Indiana builders showed a similar dramatic decline. In zinc smelting, the labor required to smelt a ton rose by 10% from 1914 to 1920. Conversely, output per man-day in Akron’s tire factories doubled from the beginning of 1920 to the middle of the recession year of 1921, and data on employment and output that Douglas himself collected from a “middle-western woodworking plant” showed an increase in output per man-hour of 35% from the first quarter of 1920 to the first quarter of 1922.

In the mid 1920s, the Bureau of Labor Statistics (BLS) began to publish systematic measures of labor productivity (defined as output per worker or per man-hour) for a large number of industries, but the time series that they constructed were short, and the steep upward
trend in labor productivity revealed by the numbers dwarfed (both on paper and in the minds of economists) any possible cyclical movements. Much attention was focused on explaining the trend, and even more on worrying over the possible consequences of the trend, with the data touching off a large literature on technological unemployment. As the economy fell into the Great Depression the BLS labor productivity indexes began to fall, but it was not clear whether this was something that typically occurred in slumps or was a special consequence of this particularly severe contraction. The bulk of the attention, however, was on whether the rapid productivity increases of the twenties might somehow have contributed to the depression and/or slowed the recovery (Woirel 2006).

During the 1930s, the BLS continued to collect the industry level data required for constructing output per worker and per man-hour measures, but the serious analysis of labor productivity indexes, both their reliability and their implications, was taking place at the National Bureau of Economic Research (NBER).¹ The research program of the NBER, from its founding in 1922 through at least the end of World War II, reflected more than anything else the vision of its cofounder and first director of research, Wesley Mitchell. Mitchell, a charismatic man and one of the most respected economists of his generation, had a vision of building an empirically based body of economic theory sufficiently reliable to serve as the basis for designing economic policies that would effectively further democratically determined social goals. At the same time, the NBER provided Mitchell with the resources necessary to pursue his own research agenda, that of creating a theoretical account of business cycles that was rooted in and tested against as

¹ Mills (1933) and Jerome (1934) report on early NBER research projects concerned with measuring and analyzing productivity statistics. In 1937 the Bureau received a grant from the Maurice and Laura Falk Foundation that was used to launch a major program on the measurement of production and productivity (NBER 1938, p. 29).
comprehensive a set of economic statistics as could be assembled. The NBER’s work with productivity statistics served both ends: evidence on longer term trends in labor productivity from a number of industries was relevant to a wide range hypotheses about the causes and consequences of technological change, among other things; while evidence on movements of labor productivity over the cycle could be used to test one of the key mechanisms in Mitchell’s original cycle theory, as described above.

Despite the interest among NBER researchers in business cycles, however, a hurdle stood in the way of coming up with reliable generalizations about the behavior of labor productivity over the cycle: The most dependable measures of output and employment came from the censuses of manufactures, which were conducted every five years up to 1919, then biennially thereafter. Annual series could be constructed from these numbers using “interpolating indexes”, and the Federal Reserve even used interpolating indexes to produce monthly measures of output per worker in manufacturing, but the careful empiricists at the Bureau were rightly nervous about deriving strong conclusions from such interpolated data. As a result, most of the analyses of productivity changes coming out of the NBER in the 1940s and 1950s dealt with long term trends.

However, NBER researchers were beginning to see patterns in the data that contradicted the conventional wisdom that labor productivity was countercyclical. Thor Hultgren, who had been assigned by Mitchell to study the cyclical behavior of the transportation sector, examined annual data from the railroads on man-hours and traffic, and concluded that “From 1921 onward, the productivity of labor, defined as traffic units per man-hour, rose in every expansion, fell in every contraction” (Hultgren 1948, p. 182). Mitchell, after seeing Hultgren’s results, wrote to his

---

2 See Biddle (1998) on the relationship between Mitchell’s conception of the social role of economic research, his own research agenda, and the program of the NBER.

3 See for example, Hultgren (1960, p. xv).
NBER colleague and coauthor Arthur Burns that “Thor virtually demolishes the notion to which I attached so much importance in 1913 that unit costs encroach upon profits late in expansion, and are materially reduced in late contraction – so far as railway transportation is concerned. He leaves mere remnants of the idea, and makes me wonder whether it has much validity in other types of business.”

George Stigler, in his 1947 NBER volume on *Trends in Output and Employment*, also noted in passing that in many industries, short run fluctuations in output were greater than those in employment (Stigler 1947, p. 46). In his posthumously published *What Happens During Business Cycles*, Wesley Mitchell acknowledged the existence of “two sets of cyclical fluctuations,” in labor productivity, “one positively, the other inversely, related to production.” The forces listed as contributing to an inverse relationship were those Mitchell had initially noted in 1913, while the main factor cited as making for a positive relationship was the view, not uncommon in the postwar years, that “modern plants attain their highest technical efficiency when operated steadily at the capacity for which they were designed.” An expansion brought actual output closer and closer to that capacity, and perhaps a little beyond it, so that measured labor productivity rose through most of the expansion. In recessions output fell below optimal capacity, dragging down labor productivity. Thus, something in the nature of the short-run production function contributed to procyclical labor productivity. Although Mitchell noted that “Fluctuations in employment as reported by the Bureau of Labor Statistics are usually slighter than the corresponding fluctuations in production”, various problems and lacunae in the available data left him unwilling to make a strong statement about whether the positive or inverse relationship dominated in the economy as a whole (Mitchell 1953, pp. 132-133).

The NBER’s official announcement of the reality of procyclical labor productivity came in the 1959 Occasional Paper *Basic Facts on Productivity Change*, with NBER Director of

---

4 Mitchell to Burns, June 16, 1946, quoted in Burns (1952, p. 53).
Research Solomon Fabricant declaring that “Over the whole period since 1889, productivity fluctuated with the state of business. Year-to-year rises in productivity were greater than the long term rate when business was generally expanding, and less, or often falling when business was generally contracting” (Fabricant 1959, p. 10). Fabricant based this fact on yet to be published research by Hultgren and John Kendrick. Hultgren had identified the most reliable monthly data on output and man-hours from mining, agriculture, and a handful of manufacturing industries, and was producing results using a statistical procedure that removed some of the trend from labor productivity, making its cyclical behavior more apparent. Kendrick was constructing measures of total factor productivity using annual data for a wider range of industries and for the aggregate economy.

Hultgren’s book, *Changes in Labor Cost during Cycles in Production and Business*, came out in 1960. Kendrick published the results of his productivity research in 1961, but gave very little attention to his findings on cyclical movements of productivity. In a forward to the book, Geoffrey Moore promised the reader that Hultgren’s results would be “highly useful” for business cycle analysis, noting that “The cyclical behavior of labor cost per unit of output . . . has heretofore been dealt with largely on a speculative basis” as the “quantitative facts on the subject” had long been “scanty” (Moore in Hultgren 1960, p. xiii). Hultgren opened by reiterating Mitchell’s 1953 discussion of the factors making for procyclical and counter-cyclical labor productivity. If the course of labor productivity over the cycle was an empirical question, however, Hultgren’s data revealed a fairly robust answer: “Hours per unit usually fall when production rises, and vice versa” (Hultgren 1960, p. 6).

---

5 Kendrick published the results of his productivity research in 1961, but gave very little attention to his findings on cyclical movements of productivity.
Hultgren’s statistical methods were those that had been developed at the NBER over the previous two decades for business cycle analysis. Using the monthly output series for each industry for which he had adequate data, Hultgren identified the peaks and troughs of each industry’s “specific cycles”. He then calculated the change in labor hours per unit of output (the reciprocal of output per man-hour) for each peak to trough period (contraction) and each trough to peak period (expansion). Pooling data across industries, Hultgren reported that

In one industry or another, at one time or another, we have data on ninety expansions of production and ninety-nine contractions. In eighty-three, or 92 percent, of the ninety expansions, there was a net decline in hours per unit. In seventy, or 71 per cent, of the ninety-nine contractions, there was a net rise in hours per unit. The pooled data suggest a strong tendency toward an inverse relation between hours per unit and total output (Hultgren 1960, p. 8).

Hultgren extended this basic analysis in several different directions. He limited the sample to industries with four or more cycles. He expanded the sample to include all industries for which reliable annual data could be found, marking peaks and trough by year instead of by month. And he examined the movement of hours per unit between the peaks and troughs that emerged when the data from all industries were aggregated into a single output series. Nothing changed the basic story of strong inverse relationship between hours per unit and output, although he did note that the inverse relationship tended to break down in the final months of expansions.

The forward to Hultgren’s book included the following footnote:

6 See Burns and Mitchell (1946) for detailed descriptions of these methods, in particular their procedure for identifying industry-specific cycle turning points. All measures were expressed relative to within cycle average values to eliminate much of the influence of secular trends.
After this paper had gone to press, the report by Edwin Kuh, “Profits, Profit Markups, and Productivity” . . . became available. This interesting analysis utilizes the national income accounts to develop quarterly estimates of output per man-hour and wage and salary costs per unit of output for the corporate sector of the economy. On most of the points that overlap, the conclusions of Kuh’s and Hultgren’s studies appear to be consistent, which is reassuring in view of their independent statistical foundation (Moore in Hultgren 1960, p. xvii).

The data used by Kuh in his report were indeed different from those used by Hultgren, and the statistical methods employed by Kuh differed from Hultgren’s as well. Kuh’s output series was the deflated value product generated by the corporate sector, quarterly, from 1947-1957. To construct a corresponding labor variable, he started with annual data on employment in the corporate sector, then interpolated quarterly employment totals using quarterly industry level employment data, with each industry’s data adjusted by the portion of that industry’s workforce employed in the corporate sector, as reported in the most recent censuses of manufactures and trade. Multiplying the employment series by quarterly data on average weekly hours from the BLS resulted in a quarterly man-hour series. Kuh’s method of determining the cyclical behavior of labor productivity was to regress output per man-hour on the level of output and a trend, explaining that “for want of a better device to measure long run technological change, the commonly employed practice of introducing a linear trend term to represent this has been followed. To measure short-run effects on output per man-hour when capital stock and techniques are assumed fixed, the level of output itself was introduced.” The coefficient on
output was positive and significant; labor productivity rose with output. However, Kuh cautioned that “This equation should only be expected to hold within a given range of output, since diminishing returns would cause a decline in output per man-hour for extremely large levels of output with given capital stock and labor force” (Kuh 1960, p. 85).

That Kuh’s choices about data and empirical methods differed from Hultgren’s reflects the fact that Kuh was part of a different research program and was motivated by different purposes. Kuh was one of a number of economists of the time working on building macroeconometric models. A central purpose of these models was to generate more accurate forecasts of key macroeconomic variables and measurements of important macroeconomic relationships, both of which would provide a more reliable foundation for economic stabilization policies. For the most part these economists worked with a Keynesian theoretical framework, and their methodological guideposts were those laid out in Haavelmo (1944) and embodied in the work of the Cowles Commission econometricians of the late 1940s and early 1950s: Economic models should be expressed as systems of non-deterministic equations and based in economic theory; the parameters of the equations should have well defined theoretical or policy relevance; and the form of the equations and the (explicitly specified) assumptions about their non-deterministic elements should allow for both the estimation of numerical values of the parameters and the application of the tools of classical statistical inference to the estimates.

Kuh was the first of several of these technocratic Keynesians who would demonstrate the existence of procyclical labor productivity in the early 1960s, including Neild (1963), who did so by using quarterly data from the UK to regress log employment on log output and a trend, and

---

7 For example, Kuh’s research on productivity movements was part of his work on the Brookings-SSRC Quarterly Model of the US Economy, and his findings were first presented in paper prepared for the Joint Economic Committee of the US Congress, a committee charged with developing policies to maintain full employment and price stability, as mandated by the Full Employment Act of 1946.
Ronald Soligo (1963) in his Yale doctoral dissertation. Work by Brechling (1965) and Eckstein and Wilson (1964) could be mentioned as well. For the most part, these economists did not articulate their motivation in terms of understanding the behavior of productivity over the cycle, but in terms of discovering or better modeling the short run relationship between employment and output, often for the purposes of producing more accurate forecasts. It became common to represent the relationship with the coefficient on current period output in an employment regression, i.e., a short-run elasticity of employment with respect to output, as Soligo (1963) termed it. These authors understood that a value below 1 for this elasticity implied procyclical labor productivity; it was just that the behavior of labor productivity was not the central focus of their work.

By the early 1960s, then, procyclical labor productivity was a well-accepted fact. For example, in his 1964 presidential address to the Econometric Society, “On the Short Run Relationship Between Employment and Output”, Solow cited the studies I have discussed above, along with a few others, and announced that “Generally speaking, productivity rises most rapidly when output is recovering toward capacity and falls or rises least rapidly during recessions. The effect appears both in British and American manufacturing statistics, but is rather more pronounced in Britain” (Solow 1964, p. 6). In 1962, Arthur Okun, while presenting to the American Statistical Association the paper that demonstrated the existence of a stable relationship between changes in the GNP and changes in the unemployment rate that came to be known as Okun’s Law, informed his audience that that “The record clearly shows that man-hour

---

8 See, e.g., Neild (1963, p. 1) or Brechling (1965, p. 211); Soligo (1963) leads off his dissertation with reference to role of macroeconomic modeling in pursuing the Full Employment Act’s mandate. Eckstein and Wilson (1964) is an exception to this generalization.
9 In the same way, Hultgren (1960) understood that the patterns he was uncovering in his labor productivity series implied that the cyclical fluctuations of output were greater than the cyclical fluctuations of employment when expressed in percentage terms; it was simply not to him the more pertinent way to express the finding.
productivity is depressed by low levels of utilization, and that periods of movement towards full employment yield considerably above average productivity gains” (Okun 1963, p. 6). And readers of Eckstein and Wilson’s (1964) paper, in circulation since 1961, were informed that Hultgren’s findings of a positive relationship between changes in output and changes in productivity at the 3-digit industry level were consistent with the characteristic behavior of productivity over the general business cycle, as had been confirmed by several other researchers using different techniques and different data.

The Labor Hoarding Concept Emerges

Okun’s 1962 paper for the American Statistical Association includes an articulation of the labor hoarding concept involving all three of the essential components that I identified in the introduction: acceptance of the fact of procyclical (measured) labor productivity, understanding of the contradiction between the fact and basic neoclassical theory, and a possible explanation of the contradiction rooted in an analysis of how fixed costs of hiring and firing affect a firm’s optimal employment strategy. It is the earliest such articulation that I have found. I have already cited Okun’s assertion of the reality of procyclical productivity. He then goes on to point out that:

The implications and explanations of this phenomenon are intriguing. Indeed, many a priori arguments have been made for the reverse view – that depressed levels of activity will stimulate productivity through pressure on management to cut costs, through a weeding out of inefficient firms and low quality workers, and through availability of more and higher quality capital per worker for those who
retain their jobs. If such effects exist, the empirical record shows that they are swamped by other forces working in other directions. (Okun 1963, p. 6)

Note here that among various a priori arguments for procyclical labor productivity, including the venerable worker quality argument, we find reference to the key mechanism that drives cyclical movements of labor productivity in the basic neoclassical theory of the firm in the short run: the quantity of capital per worker. Okun then goes on to offer some “speculation” on the matter, expressing the hope that researchers will further study the problem with “concrete evidence at a microeconomic level”:

. . .the positive relationship between output and labor productivity suggests that much of labor input is essentially a fixed cost for fairly substantial periods. Thus high output levels permit the spread of labor overheads, and low production levels raise unit fixed costs of labor. At times we may take too seriously our textbook examples which view labor as a variable factor, with only capital costs as fixed. Even the most casual empiricism points to an overhead component in labor costs. (Okun 1963, p. 6)

Okun offers many reasons why employment “may not be easily variable”, including “Transaction costs associated with laying off labor and then, in the future, doing new hiring . . .” and “Acquired skills that existing employees have learned on the job may make them particularly valuable to the firm, so that it pays to stockpile underemployed labor rather than run the risk of having to hire untrained men when business conditions improve . . .” (Okun 1963, p. 7).
Two years later, in his presidential address to the Econometric Society, Robert Solow’s recognition of the procyclicality of labor productivity was also accompanied by my two other components of the labor hoarding concept. Solow’s reference to the relationship between neoclassical theory and the short run movements of labor productivity comes in the context of a discussion of an earlier pair of papers showing that the average real wage in the economy was positively correlated with the level of aggregate output (Dunlop 1938; Tarshis 1939). These papers were intended as empirical tests of Keynes’s assertion that money wages and real wages would move in opposite directions over the cycle, and they presented fairly convincing evidence that this was not the case. But Solow recognized, as had Dunlop before him, that this test of Keynesian theory was also a test of neoclassical theory, as Keynes’s model of the demand side of the labor market in *The General Theory* was neoclassical in nature, and had implications for the cyclical movement of labor productivity as well as real wages:

. . . the system of that book, and the mathematical reformulations that followed it, contained a short-run relationship between aggregate output and employment. Whether one thinks of this relation as showing the result of applying more and more labor on an essentially homogeneous stock of capital, or the result of calling less and less efficient plants into production as output and employment increase, the effect is the same. The slope of this relation plays the role of the marginal product of labor, and the marginal product of labor diminishes as employment increases in the short run. Thus Keynes, like a good student of Marshall and Pigou, argued that any increase in employment in the business cycle had to be associated with a decrease in the real wage . . . (T)he reasoning that led Keynes to
his assertion about the cyclical behavior of real wages leads equally to the prediction that output per man must fall – or rise slower than trend – during business cycle upswings and rise – or rise faster than trend – during downswings or decelerations (Solow 1964, p. 5-6).10

Solow described three mechanisms that had been proposed to explain what he called the “perverse behavior of productivity in the short run”, explicitly giving one of them the “labor hoarding” label:

The labor hoarding mechanism operates on the assumption that important costs are associated with changes in the level of employment and with the risk that trained workers laid off as output falls may not be available as output rises again. It can be shown that a long run cost-minimization policy may require that even if labor-requirements per unit of output fall with output, employment per unit of output may well increase as recession sets in. (Solow 1964, pp. 7-8; emphasis in original).11

Solow argued that the labor hoarding mechanism was closely related to what he called the “decreasing cost” mechanism, which “rests on the hypothesis that much labor in modern industry is “overhead” in character, and can be thought of as a fixed factor in the short run.” The two mechanisms “blend into one another. Even overhead labor is not eternally fixed; many of the

---

10 Cf. Dunlop (1938, p. 431), who notes that underlying Keynes’s assertion regarding the cyclical movement of real wages is an assumption that “with given equipment and technique, an increase in employment must result in a fall in real output in per unit of labour, for industry is normally working subject to increasing cost”; an assumption that Keynes shared with the classical (i.e., Marshallian) economists.

11 Solow supports the claim “It can be shown” by citing his own paper “Distribution in the Long and Short Run . . . prepared for International Economic Association Conference on the Distribution of National Income, Sept. 1964.”
characteristics which lead employers to hoard labor are precisely those which give some labor its quality as overhead.” Solow thus chose to “use the term labor-hoarding as a catch-phrase to stand for all the frictions involved in meeting transitory variations in output with variations in employment” (Solow 1964, p. 8). In the econometric models he presented in the address, these frictions were represented by a standard partial adjustment equation for employment; and at one point he referred to an earlier version of the paper (presented in September of 1963) in which he had derived such a partial adjustment equation from a model in which employers faced fixed costs of adjusting employment and minimized “cumulated costs over a fairly long period” (Solow 1964, p. 20).12

By 1966, the modern labor hoarding concept was being taught to graduate students at Yale and presumably at other major economics graduate programs as well.13 The phrase used in its modern sense made its first appearance in a J-Stor indexed journal in 1968 when R. J. Gordon used it to explain why productivity movements in the construction industry are countercyclical, while they are procyclical in manufacturing: “In manufacturing, on the other hand, measured productivity tends to decline during business slumps, probably because labor hoarding is more prevalent than in construction, where small entrepreneurs cannot afford to pay skilled workers when there is no work to be done.” (Gordon 1968, p. 419). The nature of the usage, without definition or elaboration, shows that Gordon and the journal editor assumed that the audience would have a shared understanding of the meaning of the phrase.

12 The partial adjustment model for short term employment begins with the equation $E_t - E_{t-1} = \gamma (E^*_t - E_{t-1})$, where $E$ is employment (or its log), $E^*$ is desired employment, and $\gamma$ is the partial adjustment parameter. Then $E^*_t$ is replaced with an equation describing desired employment in terms of observable variables, which Solow derived by inverting a Cobb-Douglas production function. This leads to an estimable equation in which the change in employment is regressed on output, the capital stock, lagged employment, and a trend. In the logged version of the model, the coefficient on output is the short-run elasticity of employment with respect to output, while the long run elasticity can be calculated using the short-run elasticity and the coefficient on the lagged employment variable.

13 Personal communication with Dan Hamermesh.
Pro-Cyclical Labor Productivity and Neoclassical Theory

For Robert Solow, the newly discovered fact that labor productivity rose in expansions and fell in contractions was “perverse” because it directly contradicted a fundamental assumption of the basic neoclassical theory of the firm, the assumption of diminishing marginal productivity of homogenous labor applied to a fixed capital stock. It would be inaccurate, however, to say that Solow’s perception of the relationship between the cyclical behavior of labor productivity and economic theory was shared by all or even most of the economists working during the fifties and the early sixties on productivity measurement or estimation of the short-run relationship between employment and output.

Neither Fabricant nor Hultgren at the NBER pointed out the contradiction between the NBER’s findings regarding labor productivity and the implications of the neoclassical assumption of diminishing returns to labor in the short run, and this is not surprising. Fabricant had no particular commitment to neoclassical theory, and in this he was following in the footsteps of Wesley Mitchell, for whom the neoclassical tradition represented just one of a number of types of economic theory, each of which was potentially useful for generating interesting insights into economic activity. Mitchell had believed that the path to an accurate and socially useful economic theory involved weaving together elements derived from several theoretical frameworks in economics, but only after a careful use of statistical data to test, and if necessary modify, relevant hypotheses derived from those frameworks. As we have seen, in the early 20th century economists had a number of reasons, most unrelated to the formal neoclassical theory of the time, for believing that labor productivity was countercyclical, and Mitchell saw this a potentially important factor underlying the business cycle. This made the identification of the actual path of labor productivity over the cycle, and any corresponding movements in costs
and profit margins, a high priority activity in the NBER’s business cycle program, and it was within this context that Hultgren assessed the significance of his results.

The macroeconometric model builders working on the short-term relationship between employment and output displayed a range of attitudes regarding the theoretical implications of the patterns they were uncovering. Neild (1963) made his position towards theory clear: His first priority was to produce better forecasts for policy makers. Both “traditional” and “imperfect competition” theories, he argued, assumed that employment would fall proportionately with output; but that was not what happened in the UK. This “gap in theory” was causing problems for macroeconomic forecasters like himself. “We do not want to prove or disprove any theory”, he explained, “but to find relationships which work” (Neild 1963, p. 1). A somewhat more sophisticated version of the idea of a gap in theory was expressed in by Kuh in a 1965 paper, in which he attempted to estimate both short-run and long run elasticities of employment with respect to output by including lagged values of output in his employment regressions. In summarizing his results, Kuh commented that:

. . .the long run elasticity of demand for labor with respect to output might be less than unity – given the stock of capital – so that labor productivity will rise secularly on that account in a decidedly non-neoclassical fashion. As matters turn out, the long run parameters do not conflict with neoclassical assumption in some instances although they do in others. However, the behavior of lagged responses does indeed lead to conclusions strongly at variance with neoclassical reasoning which should be a cause of neither congratulation nor concern since the business
cycle has been excluded by choice from that domain of analysis.” (Kuh 1965, p. 9)

Thus, procyclical labor productivity was not to be regarded as a challenge to or contradiction of neoclassical theory, but as irrelevant to it.

As we shall see, not all economists of the time had made the choice that Kuh mentions, but it was not uncommon at the time for economists to accept the idea that the domain of neoclassical or Marshallian analysis was properly limited to some subset of economic activity, with alternative theoretical frameworks required to understand other, quite important, aspects of economic activity. There were various permutations of the idea. For example, one could believe that neoclassical models (the model of the perfectly competitive industry, plus imperfect competition models of the Chamberlain/Robinson variety) were the best tools for understanding the behavior of market and industry level variables, while a Keynesian approach was better for the analysis of aggregate variables. Or, one might see the domain of neoclassical theory as limited to the analysis of long run trends in those markets and industries that closely approached the competitive ideal, with other theoretical frameworks necessary for understanding short run price formation, employment determination, etc. in sectors of the economy that were imperfectly competitive, or subject to short-run increasing returns to scale. The macroeconometric model builder who believed that such sectors were quantitatively important might justify equations describing short-term aggregate relationships with appeals to theories of administered pricing inspired by Gardiner Means’ work (Means 1935), or models of full cost pricing building on Hall
and Hitch (1939). Brechling (1965, p. 188), for instance, began constructing his econometric model of the short-run employment-output relationship by assuming “imperfect competition and administered prices” and commented that he did not believe that cost minimization “or any other motivation” could be assumed over the short periods he was studying. Neild (1963), though he claimed no desire to prove or disprove any theory, opined that his results supported Hall and Hitch type models in which prices were set based on costs when operating at full capacity.

In his 1964 address, however, Solow clearly framed procyclical labor productivity as a challenge to economic theory, explicitly rejecting the attitude that “finding relationships that work” for forecasting was more important than resolving such challenges, as well as arguments that sidestepped the apparent challenge with appeals to the necessity of theoretical pluralism in economics:

I know it will be as obvious to you as it is to me that I have not solved the problem of giving a good theoretical explanation of short-run productivity behavior. I hope it will be as obvious to you as it is to me that this is a problem worth solving. Its importance goes far beyond the desirability of being able to predict how output per man-hour will change from this quarter to the next. Art Okun or Ed Kuh or Dan Suits can already do that. What I am looking for is a way to unify the economics we teach our students and the economics we use when we advise governments and analyze passing economic events, and do it in a way amenable to econometric treatment. This patching up of theory to explain

---

14 I label the full cost pricing and administered pricing models as non-neoclassical in the sense that they do not begin with an assumption of profit maximization (optimization) on the part of firms, and in some cases explicitly reject it as a behavioral assumption. See Mongin (1992) for more on full cost pricing models.
experimental or statistical uniformities is the way science usually proceeds.

(Solow 1964, pp. 29-30, skipping inserted page 29a)

Solow’s expressed sense of failure came from the fact that despite experimentation with several econometric models of employment determination, he was not able to produce one in which both the long run elasticity of employment with respect to output and the coefficient of capital were consistent with the assumption of a long run production function with neoclassical properties. That Solow considered a “good theoretical explanation” of the “perverse” behavior of labor productivity to be one consistent with the neoclassical framework is apparent from the fact that all the models he proposed in his address involved the assumption of optimization (cost minimization) and a long run production function with the standard neoclassical properties of diminishing marginal returns to labor and capital. Solow made at least two other references in the address to the desirability of unifying economic theory (pp. 13, 19), and in wishing for such a unification under a neoclassical umbrella was expressing an attitude that would soon become a dominant one in the profession.

Fixed Costs of Employment make “Labor Hoarding” an Optimal Response to Short Term Fluctuations in Demand

As we have seen, by the time Solow addressed the Econometric Society in 1964, speculation that the recently discovered procyclical labor productivity was a manifestation of various fixed or overhead costs associated with labor was prevalent. Okun (1963) had provided a list of such costs, as had Kuh (1960) and Soligo (1963). But discussions of these sorts of fixed costs, for example, which was essentially copied by Soligo, included “contractual commitments” (contractual terms of employment, severance pay); “technological factors” (indivisibilities in the production process...
or overhead costs, and assertions that they would or should influence the response of the profit
maximizing employer to variations in demand for his product, actually had a long history in
economics, going back at least to the scientific management movement of the early 20th century.

The rhetoric of the scientific management movement had a progressive, idealistic strand
that envisioned the amelioration of many of the problems facing the industrial worker via the
voluntary adoption by enlightened business owners (advised by enlightened engineers) of
personnel management policies that were rational, fair, and also beneficial to the bottom line.
Herbert Hoover’s broad approach to economic policy had its roots in this ideology (Barber
1985). In the early 1920s, then-Commerce Secretary Hoover convened a “Conference on
Unemployment,” one component of which was a gathering and impartial presentation, by the
recently founded NBER, of the important facts regarding business cycles and unemployment.
One aspect of the NBER project was research into the policies already being adopted by
successful firms to “stabilize” employment in the face of seasonal and cyclical fluctuations, as it
was hoped that wider knowledge of such policies would spur wider adoption. The quote below
comes from a chapter in the NBER study by N.I. Stone, then the General Manager of Hickey
Freeman, whose reputation as a spokesperson for enlightened capitalism would become well
established in the interwar years:

Apart from the social injury which intermittent production causes, a broad
view of the ultimate interests of the individual manufacturing concern discloses
the disadvantages of intermittent production and the gains that would flow from

that prevented certain types of labor input from being varied proportionately with the variation in output);
“transaction costs” associated with laying off labor, and then doing new hiring in the future; “acquired skills”, or
specific human capital that might be lost if the worker could not be rehire after layoff; and “morale factors”. (Okun
1963, pp. 6-7)
continuous operation . . . Business men now recognize the wastefulness of a large labor turnover, the expensiveness of training new help, and their inefficiency and resultant high cost, even when hired at comparatively low wages. The increased costs of production during the period of “tuning up” a plant to its state of efficiency, is an item familiar to every industrial engineer and plant manager. Added to these is the loss of the more capable and ambitious workers who drift away during periods of idleness to more steady occupations unless held by the inducement of higher rates of wages than would be necessary to satisfy them under steady employment . . .

These considerations have prompted a small but growing number of concerns to devise means for keeping plants in continuous operation despite the intermittent consumers’ demand (Stone 1923, p. 117).

Stone went on to describe methods employed by various firms to stabilize employment, which included labor hoarding strategies like “manufacturing to stock” and transferring “surplus help from one operation or department to another” (after instituting a policy of training workers to handle multiple tasks).

Economists were able to see these ideas discussed more fully in a neoclassical framework in J. M. Clark’s widely read book *Studies in the Economics of Overhead Cost*. J. M. Clark, son of pioneering American marginalist John Bates Clark, was one of a small number of economists of his day well versed in the marginal productivity theory of value and distribution. However, the younger Clark was less interested in further developing and refining the neoclassical model and more interested in talking about dynamic aspects of the macroeconomy, and in describing policy
implications of what we would today call market failures. So, Clark devoted much attention to
developing the idea of a social overhead cost to maintaining a healthy and properly trained labor
force, and employing the idea to highlight the social waste associated with unemployment. In the
process of doing so, however, Clark made clear that some labor costs were fixed or overhead
costs from the employer’s point of view as well. Some workers were salaried, and it was often
desirable to “keep a nucleus of the working force together through a period of depression” so as
to avoid “the cost of building up the force again when business revived” because “when new
men are taken on, there is waste in teaching them” and “there are certain key men whose
experience . . . is so valuable that they cannot be laid off without heavy loss”. Clark used
numerical examples and marginalist logic to show how these considerations affected the profit
maximizing businessman’s responses to seasonal and cyclical demand fluctuations. (Clark 1923,
pp. 43-44, 50-51, 92, 94, 160, 184-186). He concluded, however, that the marginal social cost
associated with a layoff still exceeded considerably the costs borne by the individual employer,
so that the employment reductions undertaken by businessmen in the face of reductions in
demand, while tempered by considerations of the overhead costs associated with labor, would
still exceed the socially optimal level.16

Clark did not use mathematical or graphical models in the book; his logic was expressed
verbally and he sometimes made complex points rather tersely. Still, it was a widely read book
among economists, going through seven printings over the next 25 years, and it seems likely that

16 “The effectiveness of financial self interest as a motive to regularizing production depends on how closely the
differential costs borne by the employer correspond to the true differential costs of added production to the industrial
community as a whole, and how closely the residual or constant costs of the employer correspond to the burdens
which the industrial community must bear whether production is maintained or not.” (Clark 1923, p. 167). The point
that businessmen’s self interest is not enough to lead to the socially optimal employment responses to business
cycles is found on pp. 169-170 and elsewhere.
the idea of adopting hiring policies to take account of adjustment costs remained alive among economists as well as the intellectual descendents of the scientific managers.

Interestingly, however, when the phrase labor hoarding itself began to be used by economists and business journalists to describe the business practice of maintaining a labor force larger than technically necessary, the practice was being condemned rather than lauded. During WWII, as the governments of the US and the UK were attempting to meet the demands of wartime production as efficiently as possible, they found themselves frustrated by the labor hoarding of manufacturers responding to high, but uneven, war-related demand:

Another form of manpower waste is hoarding of labor. Overstaffing may come about through managerial inefficiency or through the fear of employers that unless they obtain all the available workers at once, they will not have them when needed . . . (Joseph 1943, p. 5)

There has been, of course, a good deal of labor hoarding in certain industries, and a considerable amount of wasteful use of labor during the period when war plants were being staffed. . . . the tendency on the part of employers to take out “manpower insurance”, that is, to hire more workers than were currently needed in anticipation of future requirements, and as a protection against a high turnover rate – has been pronounced (Haber 1944, p. 61).

In the mid fifties, labor hoarding by managers in the Soviet Union was cited as a reason for higher measured unemployment in, but also the greater efficiency of, western economies (Seigel
1953, Oxenfeldt and van den Haag 1954). By the sixties, however, the pendulum was swinging back, and labor hoarding was again being seen as a sound and beneficial business practice. A. J. Brown (1958), in his presidential address to the Royal Economic Society, credited labor hoarding with keeping the British unemployment rate surprisingly low after the war. The UK’s National Joint Advisory Council of the Ministry of Labour, while admitting that inefficient “labor hoarding” might sometimes occur, concluded that:

“labor hoarding” . . . is less common than is often supposed, and that where it does occur there are often good reasons for it. Sometimes what is called “labor hoarding” may simply be the reluctance of an employer to disband a trained labor force until it is clear that there is no early prospect of his being able to employ it fully. This may well be sensible both from the point of view of manpower and from that of personnel policy. (UK Ministry of Labour, 1962, p. 47).

**Formalization**

The relatively rapid and widespread acceptance among economists of the proposition that fixed or overhead costs associated with the labor input were a likely reason for, if not the proven explanation of, procyclical labor productivity owed much to the appearance in the early 1960s of a pair of formal models characterizing the firm’s demand for labor over a multi-period time horizon. One was in the book *Planning Production, Inventories, and Work Force* by Charles Holt, Franco Modigliani, John Muth, and Herbert Simon (1960, hereafter HMMS); the other in Walter Oi’s 1961 doctoral dissertation, first published in his 1962 article “Labor as a Quasi-Fixed Factor.” Each model took the form of an abstract mathematical representation of the firm’s
decision regarding the number of workers to hire in each of several future periods, cast in the form of an optimization problem, this latter feature making both models consistent with the neoclassical paradigm. Oi assumed hiring and training costs associated with adding a new worker; HMMS described these plus various costs potentially associated with changing the size of the work force. Both models implied that changes in product demand would not be met by proportional changes in the demand for labor. Neither HMMS nor Oi, however, mentioned the implications of their model for the cyclical behavior of labor productivity.

Planning Production, Inventories, and Work Force was written while the authors were affiliated with the Graduate School of Industrial Administration (GSIA) at Carnegie Tech. The book was not presented as a contribution to the positive theory of the firm, but as a normative analysis of what the profit maximizing manager should do in the presence of uncertainty regarding future demand, with the obvious implication that while some managers may be optimizing, many are not. At the book’s core was a mathematical model described thusly: “our problem . . . (is) . . . to minimize $C_T$ subject to the relations Eqs. [4-2], with respect to the decision variables ($W_1$, $W_2$, . . . ) and ($P_1$, $P_2$, . . . ) for any given initial conditions (inventory and work force) and an arbitrary known or estimated pattern of future orders.” $C_T$ was the sum of costs over a multi-period planning horizon, eqs. [4-2] described the evolution of inventories, $W_t$ was the workforce in period $t$, and $P_t$ production in period $t$. (Holt et al., 1960, p. 93). The optimal decision rules implied by the model were derived using the calculus of variations, and readers were provided with strategies for estimating the parameters of the model’s quadratic cost function and for creating forecasts of future demand, both of which were necessary to implementing those rules. This mode of presentation was in keeping with the fact that GSIA was
a graduate business school with a mission of training managers to apply modern quantitative methods to business decisions.

It is clear, however, that the book’s model was an outgrowth of both Modigliani’s and Simon’s longstanding research goals of contributing to the positive theory of decision-making under uncertainty, and the model’s structure represented its authors’ commitment to the methodological precepts of the Cowles Commission, which called for the development of mathematically rigorous models amenable to econometric estimation. The book included a proof of an important “certainty equivalence” result: given a quadratic cost function, the optimal decision rule from a version of the model that took into account the full probability distribution of the path of future orders over the time horizon was the same as the optimal decision rule derived from a version of the model that treated the expected value of orders in each future period as the actual future value of orders. Judged by the standards of the economics profession of the time, the HMMS model and the statistical methods recommended for implementing it were quite technically demanding.

In deriving the cost function to be minimized in the selection of work force and production levels, and describing the sorts of information needed by managers to estimate it, HMMS mentioned several items that had become common in discussions of overhead or fixed costs of labor, listing and describing them under the broad headings “Hiring and Training Costs”, “Layoff Costs”, “Overtime Costs”, and “Idle time” costs. It was emphasized that there were costs associated with changing the size of the work force in addition to those fixed costs.

---

17 Holt was a Keynesian economist with an engineering background, and Muth was a graduate student. The fact that the decision model took the form of a neoclassical optimization model reflected Modigliani’s analytical preferences. The story of Simon and Modigliani’s activities and interactions leading up to the development of the HMMS model is told in Rancan (2013).

18 The authors indicated that they were simply summarizing a discussion found in a 1956 issue of the Journal of Industrial Engineering.
associated with a work forces of any given size. More important to the authors’ presentation than the precise categorization or analysis of each of these cost items, however, was the demonstration that their aggregate impact on labor cost could be approximated by a mathematical function that included linear and quadratic terms in the size of the work force, the one-period change in the size of the workforce, and the level of production (Holt et al. 1960, chapt. 3).

In the authors’ discussions of the relationship between labor costs and size of the work force, the only hint of the neoclassical concept of the diminishing marginal productivity of labor in the short run was a reference to increased costs of supervision, bottleneck delays, scrap costs, etc. when the workforce was increased in order to raise production levels. And while the authors showed that, for a given path of future orders, the implementation of the model’s optimal decisions rules would lead to a reduction in the fluctuations of both output and unemployment, the impact of these reductions on the behavior of output per worker was not discussed. (Holt et. al 1960, pp. 18-20).

Walter Oi’s 1961 approach to analyzing the impact of fixed and overhead costs associated with labor differed from that of HMMS in a number of ways. Many of these differences can be understood as consequences of the fact that Oi wrote his dissertation at the University of Chicago, where at this time students were taught a particular approach to economic analysis, one distinct from that preferred by the authors of Planning Production, Inventories, and Work Force. Some of the key tenets of this approach can be gleaned from Milton Friedman’s methodological writings of the late 1940s and early 1950s, and a number of subsequent
commentators have attempted to more fully articulate what Reder (1982) called “the Chicago view”.  

Perhaps the best known methodological precepts of this Chicago view are found in Friedman’s (1953) Essay on Positive Economics: Theories are to be judged not by the descriptive realism of their assumptions, but by the accuracy of their predictions; and that, as a consequence, economic models should be constructed so as to yield hypotheses that can potentially be empirically refuted. Beyond this, however, the Chicago view of the 1950s and 1960s included a belief that the dominant modes of economic theorizing of the day placed too much emphasis on the formal structure and logical coherence of models, and too little on their empirical application and testing. Students in the post-war Chicago department of economics learned that economic theory was a “tool for analyzing a startlingly wide range of concrete problems, rather than an abstract mathematical structure of great beauty but little power” (Friedman 1974). By economic theory was meant the neoclassical theory of competitive markets, built on the foundation provided by Alfred Marshall, whose partial equilibrium mode of analysis provided a better guide for building economic science, that is to say, expanding the explanatory reach of neoclassical theory, than the more abstract and formalistic approach exemplified by Walras’ general equilibrium theory. 

The Chicago economists’ commitment to empirical testing of theory (distinct, in their eyes, from the empirical estimation of parameters of a model presumed to be correct) plus their commitment to economic theory of a certain form, led to the development of a particular strategy

---

19 See Friedman (1953) and his earlier writings cited therein. On Friedman’s influence in shaping the Chicago view, see Reder (1982). A recent account of what the Chicago students of the 1950s were taught about economic theory is found in Hammond (2010).

20 The theory of monopoly was also a permissible analytical starting point, but monopoly was assumed to be a relatively rare phenomenon that had little impact on the overall working of the economic system. Theories of imperfect competition, even those based on the optimizing models of Robinson and Chamberlain, were rejected as methodologically unsound and of little practical use. See Friedman (1953, pp. 37-38) and Reder (1982, p. 15, 17-18).
for dealing with potential empirical puzzles or anomalies. Reder (1982), writing about the Chicago view as something of a participant observer, characterized it this way, using “TP” as an abbreviation for the Chicago approach to economic analysis:

Any apparent inconsistency of findings with the empirical implications of the theory, or report of behavior not implied by the theory, is interpreted as anomalous and requiring one of the following actions: (i) re-examination of the data to reverse the anomalous finding; (ii) redefinition and/or augmentation of the variables in the model, particularly the permissible objects of choice and resource constraints; (iii) alteration of the theory to accommodate behavior inconsistent with the postulates of rationality (constrained optimization) by one or more decision makers (resource owners); (iv) placing the finding on the research agenda as a researchable anomaly. . .

. . . Suffice it to say that TP adherents – with variation among individuals – focus attention upon (i) and (ii) and, failing quick resolution of the anomaly, move to (iv) but pay little attention to (iii). (Reder 1982, p. 13)

Echoes of this Chicago view are obvious in the opening paragraphs of Oi’s 1962 paper:

The cyclical behavior of labor markets reveals a number of puzzling features for which there are no truly satisfying explanations . . . I believe that the major impediment to rational explanations for these findings lies in the classical treatment of labor as a purely variable factor.
In this paper I propose a short-run theory of employment which rests on the premise that labor is a quasi-fixed factor. The fixed employment costs arise from investments by firms in hiring and training activities . . .

The concept of labor as a quasi-fixed factor is, in my opinion, the relevant one for a short-run theory of employment. Its implications are amenable to empirical verification and are, in the main, borne out by the available evidence. Thus my theory provides a unified explanation for various aspects of the cyclical functioning of labor markets (Oi 1962, p. 538).

In other words, there are anomalies in the cyclical behavior of labor markets for which there are no “satisfying” or “rational” explanations (meaning explanations that assume optimizing behavior and competitive markets); Oi is proposing such an explanation by adopting Reder’s strategy (ii); and his model generates empirically testable hypotheses that he will show to be consistent with empirical data. Interestingly, procyclical labor productivity was not one of the puzzles that motivated Oi.21

In Oi’s model, firms maximized profit per worker over a multi-period time horizon. There was a fixed cost to hiring or firing a worker, and Oi essentially assumed a fixed cost of training each type of worker that raised the marginal product of that type by a known amount. The firm also formed an expected value for the worker’s market wage for each future period. Solving the firm’s optimization problem led to Oi’s first important result “Even under perfect competition wages would be equated to marginal products if and only if labor is a completely variable factor”, i.e., if there were no fixed costs of hiring or training. In a long run competitive

---

21 The puzzles were occupational differences in the stability of employment and earnings, the uneven incidence of unemployment, the persistence of differential labor turnover rates, and discriminatory hiring and firing policies.
equilibrium in which neither wage nor value of marginal product for any type of labor was expected to change, each type of worker’s value of marginal product would equal his wage plus the amortized value of the cost of hiring and training him. This wedge between equilibrium wage and marginal product Oi termed the “periodic rent”, and the “degree of fixity” of a labor type was the ratio of that labor type’s periodic rent to its wage (Oi 1962, p. 541).

Oi then considered the impact on labor demand if a firm in long-run equilibrium learned with certainty of a drop in future product demand, which for a competitive firm meant a drop in future product price (P*) and a proportionate drop in value marginal product for each type of worker. Since, for workers already on the payroll, the periodic rent created by hiring and training costs was a sunk cost, the firm would keep a worker on the payroll as long as his per-period marginal product was greater than his per period wage. This led to the result that could be used to explain procyclical labor productivity: The employment of a quasi fixed factor would only be reduced if the value of marginal product, set above the wage in long run equilibrium by an amount the depended on the worker’s degree of fixity, fell below the wage. Each type of labor would have a critical P*, and only when P* fell below that level would employment of that labor type be reduced. The employment of some labor types would not fall at all with a decline in product demand. It is worth noting that while explaining this result, Oi commented that “Any decline in P* will reduce the demand for a variable factor; with falling employment, the variable factor’s marginal physical product will be increased until the equilibrium conditions are again satisfied,” thus assuming, without elaboration, a technology characterized by neoclassical short run diminishing marginal return to labor (Oi 1962, p. 541).

Oi worked through several other conjectural shocks to firms in long or short run equilibrium. He explained that his model implied that if a firm reached a short-run equilibrium
and then raised its expectation of future demand, there would not necessarily be an increase in
demand for all labor types, and there would be greater relative shifts in demand for labor types
with low degrees of fixity. He argued that if there was uncertainty in the future product price, so
that an actual $P_t$ turned out different from an earlier forecast, his conclusions about the inverse
relationship between the degree of fixity and the size of the relative labor demand shift would be
strengthened, since “the readjustment due to an error in forecasting product demand necessarily
shortens the expected period of employment, thereby increasing the magnitude of the ex-post
periodic rent.” This conclusion was not mathematically derived from the model, and as a general
matter, Oi’s approach to using his model for analysis was much less formal than that of HMMS.
For example, he never solved explicitly for an equilibrium path of future employment of a labor
type, given a future path of $P^*$. Expectation formation was not modeled; the impact of currently
held information on future expectations was not even discussed. While the construction of a
closed mathematical model in which all the key elements of the situation being analyzed were
captured, and the mathematical derivation of the key relationships of the situation from that
model, may have been a high priorities for economists identifying with the Walrasian tradition or
econometricians associated with or inspired by the Cowles commission, it was of secondary
concern for economists trained in the Chicago style of “Marshallian” neoclassicism.

Both HMMS, in their instructions to managers, and the macroeconometricians who built
short-run employment functions based on the HMMS model, used time series data to estimate
values of specific parameters of their theoretical models. Oi by contrast, identified the
correlations among observable variables implied by his model, and looked for those correlations
in the data. Using information from International Harvester on the costs of hiring and training
different classes of workers with different average wage rates, he made the case that, in general,
the average wage of a class of workers could serve as an observable empirical proxy for the
degree of fixity of that class of workers. This allowed for empirical testing of some of the
propositions of his model. For example, the hypothesis that “factors with the highest degree of
fixity and the lowest degree of substitutability with the fixed factor will experience the smallest
relative changes in employment due to any change in product demand”, became, empirically, a
test of whether high wage occupations experienced smaller rates of employment change than low
wage occupations. Oi used data from BLS industry wage structure studies done in 1928 and
1931; he was able to get average wages by occupation by state for each year from four industries.
He divided all occupations in these industries into high wage and low wage categories,
calculated the employment change over the period for each category, then conducted a chi-
square test. The results generally confirmed that the decline in employment in the high wage
occupations of an industry was smaller than the decline in employment in the low wage
occupations. Another test compared unemployment rates for various classes of workers. Since
his model implied that a shorter expected period of employment would lead to a shorter period
over which to amortize the fixed costs of employment, Oi argued that the profit maximizing firm
would want to maximize the expected future tenure of its labor force, leading to a tendency lay
off those types of employees known to be less stable. He then showed that groups known from
previous empirical studies to have longer employment tenures (middle aged vs. young or old
workers; married vs. non married workers) also had lower unemployment rates.

There was one test in Oi’s paper that could have come from the macroeconometric
literature of the time: using time series data, he estimated partial adjustment models of
employment for production and non-production workers. Since the lower-wage production
workers would have a lower degree of fixity, Oi reasoned, the estimated partial adjustment
parameter from the production worker regression should be higher than that from the non-
production worker regression, and indeed it was. Note, however, that the partial adjustment
parameter in Oi’s regressions was neither derived from nor had any counterpart in Oi’s model –
it simply measured a correlation needed to test and implication of the model.

**Stabilization of the Labor Hoarding Concept**

Although the HMMS model was presented as a normative guide for the optimizing
manager, it was picked up quickly by the leading contributors to the macroeconometric literature
dealing with the short-run behavior of employment, which, as discussed above, was one of the
venues in which the cyclical behavior of labor productivity was being discussed. Soligo (1963)
gave a verbal account of the HMMS model, and used it to justify his partial adjustment
employment equation. Kuh (1965) cited HMMS as a study of productivity movements over the
cycle. Solow (1964) did not refer to HMMS, but, as we have seen, described a similar model of
cost-minimization with labor adjustment costs from which he also derived a partial adjustment
employment equation. Thus, HMMS provided the macroeconometric research community with a
model expressed in a mathematical form that matched their methodological preferences, that
portrayed optimizing behavior, and that, when combined with a neoclassical production function,
would lead to a short run elasticity of labor with respect to output of less than unity.

And while Oi did not discuss procyclical labor productivity, the relevance of his model to
that topic was quickly recognized by economists who were concerned with the phenomenon,
being cited, for example, in Kuh (1965), Vanderkamp (1966, p. 197), and Soderstrom (1972).
This was not the major impact of Oi’s model on economic research, however. More often, it was
borrowed or elaborated upon to provide a neoclassical account of some observed labor market
phenomenon not yet so accounted for, thus following the pattern suggested by Friedman (1953, pp. 11-12): “Given that the hypothesis is consistent with the evidence at hand, its further testing involves deducing from it new facts capable of being observed but not previously known and checking these deduced facts against additional empirical evidence.” It also appears to me that many economists whose theoretical priors led them to view countercyclical labor productivity as a “puzzle” quickly accepted Oi’s model as a solution to the puzzle: countercyclical labor productivity arose from labor hoarding, which arose from optimizing behavior by firms operating in a competitive environment and facing fixed costs of hiring, firing, and training workers.

A good example of both these tendencies is Chicago-trained Sherwin Rosen’s (1969) paper offering an explanation of “The Interindustry Wage and Hours Structure”. In this paper, Rosen constructed a model of equilibrium in competitive labor markets very much in the Chicago style as described above, in which differences across industries in firm and worker characteristics would lead to different equilibrium levels of wages and hours of work. Rosen noted, citing Oi, that “The novel aspect of the model lies in the demand specification. Certainly there is a great deal of persuasive time series evidence in favor of the view that labor is not an entirely variable factor of production, since there are significant non-wage investment costs associated with it . . .”²² An indeed, the employer’s per period cost function in Rosen’s model included the term NV(q+r), where N was the number of workers, V was the fixed cost per worker, q was the retirement rate, and r the interest rate. In describing how he would convert the implications of his theoretical model into empirically testable hypotheses, Rosen observed that:

---

²² Rosen also cited HMMS, Gary Becker’s work on human capital, and the evidence presented by Brechling, Kuh, and Soligo.
Fixed labor costs involve such things as market searching costs, new employee training costs, and opportunity costs associated with on the job experience of the existing work force. These data are not available. Instead, worker attributes commonly associated with labor force quality are used as proxies. Two general observations can be offered as support for this hypothesis. One is labor hoarding during cyclical downturns. The other is occupational upgrading and downgrading during business cycles . . .If these attributes were not costly to obtain, firms would have no incentive to engage in such costly practices, and these phenomena could not be explained on the basis of rational behavior. (Rosen 1969, p. 257).

Note the structure of Rosen’s logic: labor hoarding is an observed phenomenon, the only rational (i.e. neoclassical) explanation for it is fixed costs of labor correlated with worker quality, therefore such fixed costs must exist. One can easily infer from this how Rosen would have explained procyclical labor productivity.

The HMMS model was more often used by those actively researching the cyclical behavior of labor productivity and modeling the short and long term relationship between employment and output. This work was mostly the province of macroeconometricians, and the proper specification and estimation of the labor adjustment cost function as a key structural relationship in macroeconomic models became an important research goal. Early contributors to this literature, however, who acknowledged the likely importance of the fixed cost and adjustment cost elements embodied in Oi’s model and the HMMS model, sometimes also offered an alternative explanation for procyclical labor productivity that was rooted in the technology of production rather than the costs associated with hiring labor. This second type of explanation
was related to the idea, mentioned by Mitchell in 1953, that modern plants were built to operate optimally at a particular capacity level of output, and that deviations from this level of output in either direction would lead to inefficiency. Okun (1963) had referred to “technological factors” that could lead to short-run productivity declines when output fell below “normal” levels. Solow (1964), as we have seen, distinguished “decreasing cost” explanations for procyclical labor productivity, based on “overhead” labor, from the “labor hoarding” explanations based on costs of changing the level of employment, and the language that he used in describing the decreasing cost explanations suggests that he was including explanations of procyclical labor productivity based on assumption about indivisibilities in the technology of production.

The two types of explanations – those based on the nature of the short-run production function vs. those based on the nature of the costs of hiring labor – are not mutually exclusive, as Solow noted. Arguably, however, there is a meaningful distinction to be made between the two, notwithstanding Solow’s argument about their close relationship. Optimizing behavior plus appropriate assumptions about the nature of the short-run production function could generate the labor productivity patterns uncovered by Hultgren and others even if there were no fixed costs to hiring labor, and optimizing behavior plus fixed costs of hiring labor could generate those patterns even if the short run production function was characterized by diminishing marginal productivity of labor at all relevant output levels. Only the second scenario, however, was fully consistent with the neoclassical theory of the competitive industry in the short run.

In any case, technology-based stories were offered along with labor-cost based stories in some of the important contributions to the 1960s literature on the cyclical behavior of productivity and employment. Soligo (1963), for example, (1963, pp. 19, 30-31) invoked possible indivisibilities in the capital stock “in-place”, as well as indivisibilities affecting the way
different types of labor were combined for production, as reasons that the demand for labor might not fall proportionately with demand for output. Wilson and Eckstein (1964) posited the existence of “short-run maladjustment” cost functions and labor requirement functions reflecting situations in which actual output deviates from output planned at the beginning of the production period. Ray Fair, however, in his doctoral dissertation on The Short Run Demand for Workers and Hours, surveyed the relevant literature as of 1968, and expressed his unease with authors who either assumed or concluded that the short-run production function displayed increasing returns to labor. An econometric model that attributed the positive correlation between productivity and output to increasing returns to labor was, he believed, probably mis-specified (Fair 1969, p. 31). Fair’s approach was to assume a fixed proportions short run production function that led to constant returns to labor, and then, to account for procyclical labor productivity, assume that employers usually paid for much more labor than they utilized. He justified this second assumption by pointing to Soligo’s (1963) description of the fixed costs of labor and the HMMS model, and he developed a procedure for constructing a time series index of this assumed “excess labor”.23

If we take the undergraduate economics textbooks mentioned in the introduction as a guide, explanations of procyclical labor productivity based on the characteristics of the short-run production function faded in importance over time relative to explanations based on labor costs, ratifying Fair’s opinion about the merits of the two types of explanations. Only the McConnell and Brue book (later McConnell, Brue and MacPherson) offered a production function based explanation, beginning with the observation that “firms designed their plants so that they operate with maximum efficiency during ‘normal times’”. Fair’s hypothesis about excess labor was also given something of a twist in later discussions of labor hoarding, with the suggestion that

---

23 Both Solow and Kuh were on Fair’s dissertation committee.
hoarded workers were employed in tasks such as maintenance that did not show up in the firm’s measured output. This supposition, which had long been a part of discussions of employer’s strategies for dealing with recessions, brought the idea of excess labor more into line with the basic assumption of optimization by employers, as it would certainly be sub-optimal to keep workers on the payroll doing nothing when there was something useful they could be doing. It also raised the possibility that at least some of the observed correlations between labor productivity and output was due to the mismeasurement of true labor productivity (See McConnell, Brue, and McPherson 1999, p. 545 or Fay and Medoff 1985, p. 639)

In 1985, Fay and Medoff published a paper that further bolstered the credibility of the modern labor hoarding concept. The paper (referred to in the introduction) reported on their efforts to measure directly the excess labor that Fair had assumed to exist. Fay and Medoff sent questionnaires to managers of manufacturing plants that asked, after some preliminary questions and definitions of terms, “By what percentage was the total number of blue-collar hours paid for actually below normal during the trough quarter (when real shipments were furthest below normal)?” and “If blue collar employment and hours had been cut back to the minimum technically necessary to maintain your plant’s actual level of output (and necessary operations) during the downturn, by what percentage would the number of blue-collar hours paid for have been below normal during the trough quarter?” (Fay and Medoff 1985, p. 643, emphasis in original). Managers were also asked questions about what types of work assignments were given to excess workers. The authors’ chief conclusion from their survey data was that at the trough of the business cycle, the typical plant employed 8% more labor than was technically necessary to produce its measured output, but that half of that labor was employed in other useful tasks, leaving 4% of the labor to be classified as truly “hoarded”.
Ray Fair reacted quickly to Fay and Medoff’s data, with a paper that actually appeared in print six months prior to the publication of the Fay and Medoff paper. Fair’s goal was to apply his own indirect method to measuring excess labor to aggregate data, “to see if the quantitative estimates of Medoff and Fay are consistent with the aggregate evidence.” Fair went on:

If this is the case, which the results in this paper show, it provides a strong argument in favor of the excess labor hypothesis. Essentially the same conclusion has been reached using two very different data sets. This is one of the few examples in macroeconomics where a hypothesis has been so strongly confirmed using detailed micro data” (Fair 1985, p. 239).

Conclusion

By the 1990s, the modern labor hoarding concept had come to be widely accepted by economists as the most likely explanation of the observed procyclicality of labor productivity, an explanation that was based on the assumption of profit maximization by the firm, and was consistent with the neoclassical model of a competitive industry in short run equilibrium. As far as many economists were concerned, models of labor as a quasi-fixed factor had successfully expanded the explanatory power of neoclassical theory, and assimilated a potential anomaly.

Interestingly, however, the researchers who first produced the convincing evidence that labor productivity was procyclical were not engaged in an attempt to test the basic neoclassical theory of production. The perception of the finding as an important puzzle or paradox for neoclassical theory was in large part due to Robert Solow’s decision to frame it that way some
three years after it first became widely known. The idea that was at the heart of the resolution to the paradox – that fixed costs associated with hiring, firing, and training workers would lead profit maximizing employers to retain more labor than was technically necessary during cyclical downturns in product demand – had been discussed by economists and others for decades as a shrewd business strategy that might also have some social benefits, or, under some circumstances, social costs. About the time that the profession was coming to believe that labor productivity was procyclical this idea was formalized in two influential models, but neither of these two modeling efforts was motivated by a desire to explain the cyclical behavior of labor productivity.

In the introduction I shared my suspicion that most economists familiar with the modern labor hoarding concept probably believe that the sequence of events leading to its emergence as a plausible explanation of the behavior of labor productivity over the cycle was more orderly than this paper has shown it to have been. I also suspect that case studies of the development of other widely accepted explanations for empirical findings initially perceived by economists as puzzles would reveal similarly complex histories.
References


__________.*. *The Short Run Demand for Workers and Hours*. Amsterdam: North Holland, 1969.


