ECONOMIC REVIEW

Economic Review is published quarterly by the Research Department of the Federal Reserve Bank of Cleveland. Copies of the *Review* are available through our Public Information Department, 216/579-2157.

Coordinating Economist:
Randall W. Eberts

Editor: William G. Murmann
Assistant Editor: Robin Ratliff
Design: Michael Galka
Typesetting: Liz Hanna

ISSN 0013-0281

Can Competition Among Local Governments Constrain Government Spending?

by Randall W. Eberts and
Timothy J. Gronberg

Randall W. Eberts is an assistant vice president and economist at the Federal Reserve Bank of Cleveland. Timothy J. Gronberg is an associate professor of economics at Texas A&M University.

The authors wish to thank Ralph Day and David D'Alessandris for excellent computer assistance and Erica Groshen and Daniel McMillen for helpful comments and suggestions.

Introduction

The United States contains more than 80,000 separate governmental units. If none of these units overlapped, each government would serve fewer than 2,000 individuals. Governmental units do overlap, however, resulting in several layers of jurisdictions. Residents within a metropolitan area typically receive public services from a municipality, a township, a county, and a host of special districts.

In addition, at each level of government, several similar governmental units may provide services within the same geographical area. For example, the Chicago metropolitan area alone contains more than 250 municipalities, each responsible for the same array of governmental functions. Overlapping these governments are 835 special districts, which usually perform only a single function, such as providing regional transportation or enforcing environmental protection regulations.

The impact of this structure on government behavior is varied, and the net effects are not yet fully understood. Critics of the decentralized structure of local governments blame the proliferation of local governments for what they see to be "runaway" spending. They argue that duplication of efforts by similar but independent jurisdictions within the same geographical area is an inefficient way to provide public services and that the resulting fragmentation could negate any benefits derived from economies of scale.

Proponents of a decentralized public sector counter with the argument that it fosters increased efficiency in the production of public goods. They maintain that competitive pressures induce local governments to adopt the most efficient provision techniques and to tailor the levels of provision of public goods to the preferences of societal subgroups (Oates [1972]).

The phenomenal expansion of the local public sector adds fuel to this controversy. Since 1950, state and local government expenditures have increased at a faster rate than either the gross national product, federal expenditures, or expenditures on private-sector services. State and local governments currently claim 17 percent of total personal income, in contrast to 10 percent in 1950. Currently, they spend two and one-half times more than the federal government spends on civilian services such as education, roads, welfare, public health, hospitals, police, and sanitation.

How much of this growth is due to government structure and how much is due to other factors, such as demand for local services, is an empirical question. Even the effect of governmental structure can work in opposite direc-

tions. For instance, a decentralized public sector may increase local public spending due to duplication of efforts, but at the same time, competition among these units may constrain spending. The net effect of our present governmental structure on government spending depends on which of these various factors is more important.

To further complicate matters, there are two distinct types of local governments. One type provides a variety of services to a subgroup of the county or metropolitan population, while the second type typically provides a single service to the entire local area. Possible differences in behavior of these two government types must be taken into account. Two previous studies, one by Oates (1985) and a follow-up by Nelson (1987), have estimated the relationship between decentralization and government spending, but without conclusive results.[1]

The purpose of this paper is to continue the inquiry into the relationship between decentralization and the size of the local public sector. We test the decentralization hypothesis proposed by Oates, in which an increase in the number of governmental units reduces local government spending as a percentage of personal income. However, unlike Oates (and Nelson), we contend that the hypothesized effects will most likely be observed at the metropolitan and county levels (referred to as the local level), not at the state or national levels. We believe that most of the "discipline" derived from competition for households and firms would be observed at these levels of disaggregation, because these levels more closely approximate local labor markets within which firms and labor are most mobile. Oates (1985), in fact, argues that the "discipline" resulting from fiscal competition should increase as the geographical size of the unit of analysis decreases. However, neither Oates nor Nelson uses a unit of analysis less aggregated than the state.

To test our point, we use various levels of aggregation from the county to the state level. We find solid statistical support for the decentralization hypothesis at the metropolitan and county levels. Increases in the number of competing general-purpose government units are associated with a statistically significant decrease in the relative income share of local public expenditures. At the same time, we find a distinct difference in behavior between the two types of government. An increase in the number of single-purpose districts increases the share of personal income going to local government expenditures. To further support our point, we find that these relationships are not significant at the state level, which is consistent with the results of Oates and Nelson.

I. Competition Among Local Government Jurisdictions

The potential benefits of competition among local government jurisdictions are similar to the benefits associated with competition in private markets. In the private sector, competition induces profit-maximizing firms to provide goods or services preferred by consumers at the lowest resource cost. The motivating force behind this behavior is the choice of suppliers available to consumers. If a firm raises its price, consumers will switch to the supplier with the lowest price, assuming that all firms are identical and that consumers incur no additional cost in searching for another supplier. Given enough competing firms (that is, choices to the consumer), no firm can set prices above the per-unit cost of production.

The same competitive forces exist among local government jurisdictions. By law, local governments cannot earn profits. However, according to Niskanen (1971), public administrators may be motivated to maximize revenue, and thus expenditures, in order to expand desirable aspects of their working environment. Public administrators thereby "consume" profits on the job instead of taking them home.

The capacity of governments to increase revenues depends upon the customer base—taxpayers who live within their jurisdictions. If local governments attempt to raise taxes or to reduce the level and quality of services, then taxpayers will have an incentive to locate in neighboring jurisdictions that provide a service/tax package more in line with the taxpayers' preferences. The loss of households and firms reduces a government's tax base and, in turn, reduces its ability to raise revenue.

Thus, the basis for the constraining effect of decentralization is founded upon the inter-jurisdictional competition for mobile resources, both labor and firms. The line of argument

■ **1** An unpublished paper by Zax (1987), recently brought to our attention, also takes exception to the use of state-level data by Oates and Nelson. Using county-level data, he finds a negative and statistically significant relationship between the number of governments and the size of the local public sector. His study differs from ours in at least three ways. First, he uses own-source revenue as a dependent variable, whereas we use local expenditures on selected functions. Second, we explore these effects at various levels of aggregation, not just at the county level. Third, he finds that an increase in the number of special districts also reduces the size of the local public sector. We find the opposite effect at each level of disaggregation.

follows the old industrial-organization paradigm of structure, conduct, and performance. Applied to the public sector, the argument runs from an increase in the number of independent public jurisdictions (suppliers), to an increase in the degree of competition, to a decrease in the relative size of the public sector. However, the efficacy of governmental fragmentation depends on the mobility of households and firms.

The net benefit of the move determines the extent to which mobility occurs or is likely to occur. This benefit comes from either the savings derived from locating in a lower-cost jurisdiction or the advantages gained from residing within a jurisdiction that provides more or better services, everything else being equal.

The costs associated with choosing between local governments are generally greater than the costs incurred in searching for alternative suppliers of private goods and services. To change local governments, a household must change residence and incur the costs of purchasing a new home and finding a new job, and must bear the emotional costs of moving to a new area.

However, these costs are in direct proportion to the distance one must move in order to find a more preferable governmental unit. For example, if enough choices of local governments are available within the same metropolitan area, then the discontented taxpayer may not need to change jobs in order to change jurisdictions. Consequently, the mobility of households and firms increases as the size of the geographical area decreases. Therefore, we would expect local governments to be more constrained by competitive forces at the county or metropolitan level than at the state or national level.

The two empirical studies by Oates and Nelson have looked for the constraining effect of competing jurisdictions only at the state level. Oates proposes and tests the hypothesis that the size of the public sector should vary inversely with the extent of fiscal decentralization, other things being equal. He uses the number of jurisdictions within each state as a measure of decentralization. Using state-level aggregates, however, he finds no significant relationship between state and local expenditures as a percent of state personal income and the number of jurisdictions.

In a reply to Oates' paper, Nelson suggests two modifications. The first is to distinguish between general-purpose jurisdictions (such as municipalities) and single-purpose jurisdictions (such as school districts and mosquito-abatement districts). Nelson argues, and rightfully so, that the two types of districts are not comparable and consequently should not be lumped together. The multiplicity of special districts within a metropolitan area does not necessarily indicate that consumers have a choice, but rather that residents are provided several services, each by a different district.

In addition, since many special districts provide only minor services and since nearly half of them lack the authority to levy taxes, Nelson argues that there may be little incentive for individuals to choose between these districts. The second modification is to include state-mandated programs in the analysis to account in some way for differences in functional responsibilities among jurisdictions. With these modifications, Nelson finds the desired systematic relationships, but the precision of the estimates is below the usual acceptable confidence level.[2]

II. Market Structure of Local Governments

As mentioned previously, one of the prerequisites for competition is a sufficient menu of choices offered to consumers. Tallying up the number of local governments in the United States casts little doubt on the potential for choice. According to Aronson and Hilley (1986), 79,862 governmental units below the state level existed in 1977. These units tend to fall into two categories: general-purpose and single-purpose governments.

General-purpose governments, such as municipalities and counties, provide a variety of services ranging from fire protection to health care. As shown in table 1, municipalities numbered more than 18,000 in 1977, or 24 percent of all governmental units; counties totalled 3,042, or less than 4 percent. Single-purpose units, consisting primarily of school districts and special districts, comprise the majority of local government jurisdictions. As noted in table 1, over 40,000 governmental units have been established to provide only a single function. More than half of these units are special districts, which include sanitary districts, drainage districts, and soil-conservation districts.

■ **2** Nelson does find the desired statistically significant relationship between the number of general-purpose governments and the size of the local public sector using state-level data. However, in what we take as Nelson's most preferred specification, equation (3) and dependent variable G*, the coefficient on the general-purpose-government variable has a t-value of only 0.91. Thus, although we are in total agreement with Nelson's methodological changes, we do not believe that a clear vindication of the decentralization claims utilizing the state sample has been established.

The overlapping structure of local governments is far from static. Between 1957 and 1977, the number of local governments fell by 22,514, primarily from a conscious attempt to consolidate local school districts. The reduction in the total number of units would have been much

TABLE 1

Type of Government	Number of Units			
	1957	1967	1977	1982
County	3,047	3,049	3,042	3,041
Municipality	17,183	18,048	18,862	19,076
Township and town	17,198	17,105	16,822	16,734
School district	50,446	21,782	15,174	14,851
Special district	14,405	21,264	25,962	28,588
Total	102,279	81,248	79,862	82,290

SOURCE: Numbers obtained from Aronson and Hilley (1986), Table 4-1, p. 76.

greater during this time if it were not for the creation of more than 11,000 special districts. Between 1977 and 1982, the proliferation of special districts continued, while the number of other types of governmental units remained relatively constant.

As expected, local governmental units are concentrated in metropolitan areas. We find that counties in Standard Metropolitan Statistical Areas (SMSAs) have almost twice as many governmental units as do non-SMSA counties—an average of 40 compared to 21. The ratio is even higher for single-purpose units (2.3 to 1), but it is smaller for general-purpose governments (1.6 to 1). In addition, we find that only 25 percent of the metropolitan areas had fewer than 10 general-purpose units and 14 single-purpose districts. On the other hand, 50 percent of the SMSAs contained more than 21 general-purpose units and 29 single-purpose districts.

III. The Empirical Test

The basic relationship to be tested is between government performance and market structure. The specification and analysis in this section follow the lines initiated by Oates and Nelson. The principal difference in our study is that we focus solely on local government expenditures in local labor markets, rather than on the aggregate of the state and local public-goods sectors. Consistent with this focus, we adopt two levels of aggregation as the geographical unit of observation: the county and the metropolitan area. In addition, as a point of reference to the previous two studies, we also estimate the relationship at the state level.

Our data set consists of observations on local public-sector characteristics and relevant demographic features of more than 2,900 counties and 280 SMSAs in 1977. This year was chosen for two reasons. First, it is consistent with the studies by Oates and Nelson. Second, some information, such as state mandates, was available only during this period. We have analyzed more current data on local-government expenditures for 1985, while still using state mandates from the earlier period, and find no qualitative differences in the results.

Variables

Local government performance is measured by expenditures on the major local public services as a percentage of personal income in either the county or the SMSA, whichever is appropriate. We include local expenditures on local schools, public welfare, fire and police protection, sanitation, and local parks.[3]

The key explanatory variable is market structure, which is measured by the number of local governments within the appropriate unit of observation. Local governments are divided into the two classes described earlier: general-purpose and single-purpose jurisdictions.[4] Three different measures of the number of local governments are used in the analysis. The first measure is simply the total number of each class of local governmental units found within the appropriate unit of analysis (county or metropolitan area). The second method normalizes the number of units by the size of the population served by all of these local governments. The third method divides the number of juris-

■ **3** Nelson did not include police protection in his estimation. We find, however, that the results are not sensitive to its inclusion or exclusion.

■ **4** The number of general-purpose governments is the sum of the number of county and municipal governments, except in Pennsylvania, New Jersey, and the New England states, where townships are also included. The number of single-purpose governments is the sum of the number of townships, school districts, and special districts, except in the aforementioned states, where townships are not included. The reason for the exceptions is that the functional responsibilities closely resemble municipalities in these states.

dictions by the total land area in the county or SMSA. This last method accounts to some degree for the ease of mobility among the various governmental units.

The other explanatory variables include state mandates, per-capita personal income, population, and intergovernmental grants as a percentage of total local tax revenues. The first three variables may be considered proxies for the demand for local public services. As Nelson notes, state mandates may impose binding minimum constraints on certain local government activities. As defined by the Advisory Commission on Intergovernmental Relations (ACIR), which collected the data, a state mandate is a legal requirement imposed by the state that a local government must undertake a specified activity or provide a service that meets minimum state standards.[5] The presence of such restrictions would, therefore, be positively associated with the relative size of the local public sector.

The demand for local public services should be positively related to personal income, according to traditional consumer demand theory. However, the relationship between *per capita* income and government spending as a percentage of personal income has been subjected to considerable empirical scrutiny. Investigation of Wagner's "law" or, perhaps more correctly, Wagner's hypothesis of a positive correlation between income and government's relative claims on that income, has sparked much research and has kindled considerable controversy.[6] To our knowledge, the empirical studies have all involved national samples. Our study will provide a simple test of Wagner's "law" at the local level.

An increase in population, holding other variables constant, would also be associated with a larger local public sector. This result in some ways follows the thinking of Wagner, who saw an increase in population density and urbanization leading to increased public expenditures on personal protection and economic regulation (Bird [1971]).

The ratio of intergovernmental grants to local tax revenues measures the extent to which local governments rely on higher-level governments for funds. Because of the matching provisions of many federal and state grants, we would expect the grants to stimulate local government expenditures.[7]

Results

Fourteen separate models were estimated: one for each level of aggregation and for each measure of decentralization. The estimates displayed in table 2 for one of the models are typical of the results found for the other models. We find that an increase in decentralization of *general-purpose* governments, measured by any one of the three measures, is statistically significantly related to a decrease in the size of the local public sector. This finding supports the decentralization hypothesis: an increase in jurisdictional fragmentation is associated with a decrease in local budget share.

On the other hand, we find that an increase in the number of *single-purpose* units increases the local budget share. This suggests that the costs of providing services through special districts outweigh the constraining effects that competition may impose on spending or the savings that result from economies of scale. Thus, our results support the argument that the proliferation of special districts has increased local spending.

The negative and significant coefficient on per capita income is evidence against the relevance of Wagner's hypothesis applied to the local government sector. At the state level, we find a positive relationship, as does Oates. A negative correlation between local public-expenditure share and income is not unexpected, however. Most studies of local public-expenditure demand find income elasticities that are significantly less than unity, which implies a decline in aggregate budget share as average community income rises.[8]

The positive coefficients on the population and intergovernmental transfer variables are consistent with our earlier discussion.

■ **5** The ACIR surveyed local governments about 77 functional subcomponents in five broad areas: state personnel, other than police, fire, and education (15 components); public safety (31); environmental protection (8); social services and miscellaneous (10); and education (13).

■ **6** Bennett and Johnson (1980) provide a comprehensive summary of the debate and a compendium of the empirical results. Ram (1987) appears to have made the most recent contribution to the literature.

■ **7** King (1984) offers a comprehensive summary and critique of the effects of grants on local government spending.

■ **8** Inman (1979) includes a summary of studies of the demand for local public services.

TABLE 2

Variables	Mean (Standard error)	Coefficient (T-statistic)
Number of general-purpose units	28.8 (40.83)	−.015 (4.48)
Number of single-purpose units	54.1 (80.55)	.005 (2.79)
Per capita income ($1,000s)	6.67 (.98)	−.317 (2.87)
Ratio of transfers to local taxes	1.18 (.53)	.559 (3.02)
Population in SMSA (100,000s)	5.53 (10.04)	.45 (3.85)
Total state mandates	37.0 (11.92)	.083 (11.59)
Constant		5.23 (6.17)
Dependent variable: local expenditures per personal income	6.94 (1.80)	
Number of observations	289	
R-square	.43	

SOURCE: Government expenditure data from Census of Governments, 1977; personal income and population data from the Bureau of Economic Analysis; state mandates compiled by the ACIR.

Various Measures of Decentralization

The conclusion that increased decentralization of general-purpose governments is associated with a smaller local public sector is supported by our analysis regardless of which measure of decentralization is used. As seen in table 3, not only are the coefficients statistically significant at the 1 percent level for SMSAs and counties, but the magnitudes of the elasticities are also of similar magnitudes, with few exceptions. For example, at the SMSA level (column 1), we find that a 10 percent increase in the number of general-purpose jurisdictions reduces the local public sector's share of personal income by 0.6 percent. In the case of SMSAs, a 10 percent increase in general-purpose governments would mean only an additional three units.

However, when state-level data are used, the statistical significance of the estimates falls below the 10 percent confidence level. The only exception is the effect of the number of general-purpose governments, which is statistically significant right at the 10 percent level.

Table 3 also reveals that the size of the local public sector at the SMSA level is slightly more responsive to a change in the number of general-purpose governments than to a change in the number of single-purpose governments. This relationship holds no matter which decentralization measure is used, but is less consistent at the county level.

IV. Conclusion

We have found a significant relationship between governmental structure and government size. Two basic relationships emerge from the analysis. First, an increase in the number of general-purpose government units within a metropolitan area or county boundary reduces the share of personal income going to the local public sector. Second, an increase in single-purpose government units has the opposite and equally significant result of increasing the size of the local public sector.

The difference in behavior between the two types of governments underscores our conclusion that competition among local general-purpose governments constrains local government spending. Recall that suppliers are disciplined by the presence of other suppliers only when they provide similar services to the same market. General-purpose governments meet this requirement more closely than do single-purpose governments. Typically, a single-purpose government is the sole supplier of a specific service within a local market, whereas each general-purpose district provides a similar array of services.

Thus, the existing structure of government creates two opposing forces of government behavior. Competition among general-purpose units, such as municipalities, constrains local government spending. On the other hand, the overlapping labyrinth of single-purpose governments stimulates local government spending.

Much of the current arrangement of local governments resulted from attempts by states and localities to respond to changing conditions within the various constraints imposed on them. As a practical matter, states and municipalities have limited ability to respond to changing conditions. States are constrained by local loyalties, vested interests, and the inertia of the

TABLE 3

Measure of Competition	Level of Aggregation				
	SMSA	County			State
		All	Non-Metro	Metro	
A. Number of units					
General-purpose	−.063	−.045	−.043	−.054	−.069*
Single-purpose	.040	.034	.046	.042	.005**
B. Number of units per capita					
General-purpose	−.076	−.036	−.062	−.068	−.032**
Single-purpose	.050	.035	.045	.033	.019**
C. Number of units per square mile					
General-purpose	−.065	−.018	−.022	−.016	—
Single-purpose	.055	.005	−.028	.023	—

Note: Numbers are expressed as elasticities. All estimates are significant at the 1 percent level unless denoted by an asterisk. A single asterisk denotes significance at the 10 percent level but less than 5 percent level. A double asterisk denotes significance at less than the 10 percent level. The estimates are derived by regressing the local government expenditures as a percent of personal income against measures of government competition, population, per capita income, intergovernmental revenue, and state program mandates. Estimates of a typical regression equation are shown in table 2.
SOURCE: Authors.

status quo. The power of localities to handle public services is often made difficult by state statutes that limit powers to tax and to incur debt.

Since the late 1950s, special districts have been established as a means of circumventing these constraints by shifting responsibilities away from general governments. The federal government has further stimulated the creation of special districts through "direct advocacy." Many federal agencies would rather deal directly with officials of special districts than with officials from general governments such as counties or municipalities (Aronson and Hilley [1986]).

In the past few years, a number of states have begun to take a systematic look at the current structure of local governments. Several states have established advisory commissions to consider reorganizing and streamlining the perceived fragmented system of local governments that dot their landscape. These commissions appear to be particularly concerned about how the large number of special districts affects the provision of services.

Our analysis provides some information that may be useful to these reform efforts. First, our results suggest that reform efforts directed toward special districts are well-guided. Clearly, an increase in the number of single-purpose governments, which consist mostly of special districts, increases government spending. Although these results are very strong, we should caution that we have not been able to control entirely for differences in the level of services provided by these governments. It may be the case that part of the observed increase in spending associated with greater numbers of units simply indicates that additional special districts are providing additional services.

Second, our results warn against lumping together general-purpose and single-purpose governments when considering streamlining local government structure. We show that the two different types of governments exhibit distinctly opposite behavior.

Third, our results suggest that a competitive environment among specific types of local governments can constrain government spending and promote the efficient provision of local public services.

REFERENCES

Advisory Commission on Intergovernmental Relations. *State Mandating of Local Expenditures*. Washington, D.C.: U.S. Government Printing Office, 1978.

Aronson, J. Richard, and John L. Hilley. *Financing State and Local Governments* (4th ed.). Washington, D.C.: The Brookings Institution, 1986.

Bennett, James T., and Manuel H. Johnson. *The Political Economy of Federal Government Growth: 1959-1978*. College Station, TX: Texas A&M University, 1980.

Bird, Richard M. "Wagner's 'Law' of Expanding State Activity." *Public Finance*. 26 (1971): 1-26.

Inman, Robert P. "The Fiscal Performance of Local Governments: An Interpretive Review." In Mahlon Straszheim and Peter Mieszkowski, eds., *Current Issues in Urban Economics*. Baltimore, MD: Johns Hopkins University Press, 1979: 270-321.

King, David N. *Fiscal Tiers: The Economics of Multi-Level Governments*. London: Allen and Unwin, 1984.

Nelson, Michael A. "Searching for Leviathan: Comment and Extension." *American Economic Review*. 77 (March 1987): 198-204.

Niskanen, William A. *Bureaucracy and Representative Government*. Chicago: Aldine-Atherton, 1971.

Oates, Wallace E. *Fiscal Federalism*. New York: Harcourt Brace Jovanovich, Inc., 1972.

_____. "Searching for Leviathan: An Empirical Study." *American Economic Review*. 75 (September 1985): 748-57.

Ram, Rati. "Wagner's Hypothesis in Time-Series and Cross-Section Perspectives: Evidence from 'Real' Data for 115 Countries." *Review of Economics and Statistics*. 69 (May 1987): 194-204.

Zax, Jeffrey S. "Is There a Leviathan in Your Neighborhood?" New York: City University of New York, Department of Economics, November 1987.

Exit Barriers in the Steel Industry

by Mary E. Deily

Mary E. Deily is a visiting economist at the Federal Reserve Bank of Cleveland and an assistant professor of economics at Texas A&M University. The author would like to thank Paul Bauer, Randall Eberts, Erica Groshen, Stephen Karlson, and Mark Sniderman for helpful comments.

Introduction

The U.S. steel industry seems perennially afflicted with overcapacity. Even after numerous plant closings, and despite recent high capacity-utilization rates, analysts suggest that another 15 to 20 percent of current capacity should close. Why has overcapacity been a chronic ailment of steel firms during the 1970s and 1980s? Why haven't firms closed plants more quickly, since continued operation of these plants depresses profits for the entire industry?

The persistent survival of excess capacity is not inexplicable. In theory, a market system reallocates resources from activities yielding lower-than-normal returns to activities with higher returns. In practice, however, firms can be locked into a low-profit activity if large losses are incurred when capital is transferred to new activities. These potential losses form an exit barrier, delaying plant closings, depressing profits, and prolonging adjustment for the entire industry.[1]

The primary purpose of this paper is to examine the nature and size of exit barriers in the steel industry. First, the necessity for contraction in this industry is summarized. Then basic exit theory is reviewed, and several types of exit barriers that seem most pertinent to the steel industry are described. The potential size of these barriers in the steel industry is assessed. Finally, the possible effects of current trade-protection and pension-insurance policies on the size of exit barriers in the steel industry are discussed.

This paper argues that high exit barriers have significantly slowed the industry's contraction by delaying plant closings. These barriers explain why capacity has fallen slowly even though industry profits have been subnormal since the late 1950s. They also help to explain why the industry failed to modernize some plants, even though these increasingly inefficient plants continued to operate into the 1980s.

I. The Necessity for Contraction

The U.S. steel industry has performed poorly during the last 25 years. Profits for the industry have been low compared to the average manufacturing return in virtually every year since

■ 1 The term "exit barrier" is perhaps unfortunate, as it carries the connotations of inefficiency attached to the phrase "entry barrier." Such is not the case: exit barriers are the various cost conditions that make lengthy exit a rational response by firms.

1958.[2] And despite the industry's recent buoyant performance—part of which appears to be due to trade protection—long-run trends in steel demand and steel supply point to continued low profits in the future.

Structural changes in steel demand have greatly reduced the growth of the market. These changes, which include increased use of steel substitutes such as aluminum and plastic, and reductions in the amount of steel used in consumer durables, particularly cars, have reduced the U.S. economy's need for steel. The average annual growth rate of U.S. apparent steel consumption has fallen from from 4.1 percent during 1960-1969, to 1.9 percent during 1970-1979, to 0.2 percent during 1980-1986.

Not all steel firms have fared the same, however. The industry basically consists of two parts: integrated mills and minimills. The integrated mills, which produce steel from iron ore, are the traditional steel industry, while the minimills, which produce steel products by recycling steel scrap, are relative newcomers. It is the integrated portion of the industry that has performed so poorly; minimills have flourished, increasing their market share from about 3 percent in 1960 to 18 percent in 1985.

As their name suggests, minimills produce steel on a much smaller scale than integrated plants, reducing the size of the required capital commitment considerably.[3] The mills also benefit from employing workers at lower wages. Though their costs are extremely sensitive to the price of scrap, minimills have become very competitive in the product lines in which they specialize, drastically reducing the integrated mills' sales in these markets.[4]

In addition, integrated firms in the U.S. faced tough new competition from imports for a share of the market, as fundamental changes in input costs during the 1950s and 1960s altered the comparative advantage in steelmaking. Two studies, by Crandall (1981) and by Kawahito (1972), examine the changes in the relative cost of materials in the U.S. compared to other countries, particularly Japan. Formerly, abundant supplies of coal and iron ore assured U.S. producers of a materials cost advantage that, along with greater U.S. productivity, more than compensated for higher U.S. wage rates. However, the discovery of rich iron-ore sources in several parts of the world and the decreased cost of ocean shipping began to reduce the traditional U.S. advantage.

Also, as Barnett and Schorsch (1983) point out, countries like Japan experienced phenomenal growth in steel consumption after World War II. Their steel industries were able to build entirely new, large-scale plants, since their rapidly expanding markets could easily absorb the output of the additional capacity. These new plants incorporated the latest technology into an optimal plant layout, resulting in high productivity growth. Increased productivity growth, combined with lower wage rates, reduced the unit cost of labor further below U.S. levels. This advantage, added to the favorable changes in materials costs, made foreign steel very competitive with U.S. integrated production.[5]

The result has been a decline in the market share of integrated steel firms in the U.S. from more than 90 percent in 1960 to less than 65 percent in the 1980s. Given the slow growth of the market, these figures translated into a need to cut integrated steel capacity by closing plants. And, in fact, the industry has closed plants. From its height in the early 1970s of approximately 155 million tons, annual raw steel capacity has fallen to about 112 million tons.

But the contraction of the industry has taken a long time, even though capital has been earning subnormal profits for many years.[6] Rather than moving into other activities, firms appear to be clinging tenaciously to capacity by nursing along aging plants, as if the growth in demand for steel might miraculously increase to pre-1970 levels. But as the discussion in the next section shows, this response may well be optimal for firms facing high exit barriers.

■ **2** See Crandall (1981), p. 29, for the rate of return on equity after taxes in steel versus all U.S. manufacturing for the years 1954-1978. See U.S. Department of Commerce, Bureau of the Census, *Quarterly Financial Reports for Manufacturing Corporations*, various issues, for subsequent years.

■ **3** Minimills typically consist of an electric steel furnace, a continuous billet caster, and some kind of finishing mill, usually for bars. See Miller (1984) for a good description of this technology.

■ **4** Minimills have a cost advantage over all integrated mills, whether domestic or foreign, in the products they can produce. See Barnett and Crandall (1986) for a detailed comparison of minimill to integrated mill production costs.

■ **5** In fact, Crandall (1981) concludes that a totally new integrated plant would be a poor investment in the United States, given his estimates of the possible reductions in labor and energy savings attainable.

■ **6** The first major plant closings, those of Youngstown Sheet & Tube and the United States Steel Corporation at Youngstown, did not occur until the late 1970s, and the next episode did not occur until 1982. In addition, because capacity is usually measured as the ability to produce raw steel, estimates of capacity reductions may be somewhat overstated. The introduction of continuous casters has increased the yield from raw steel by 10 to 15 percent.

II. A Model of the Plant Closing Decision

The neoclassical prediction for a competitive industry facing an inward-shifting demand curve is that high-cost plants will exit, leaving the lowest-cost plants to produce in the long run. However, as long as variable costs are covered, a firm will continue to operate an exiting plant that has fixed costs, since doing so minimizes the firm's losses.[7] During this period the firm will not make any major reinvestments; instead, it will disinvest from the capital in place.

Because most production processes do involve fixed costs, the decision to close a plant usually will involve a period of operation and disinvestment before shutdown. The optimal closing point will not occur until the net revenue, which is the return to continued operation of the capital in place, equals the return that could be earned on the salvage value. Thus, the speed with which a firm closes a plant depends on how quickly net revenues decline and on the amount of capital that can be salvaged once the plant is shut down.

Clearly, one important factor that will affect the timing of plant closings is the general level of economic activity. When sales decline during recessions, they increase the probability of plant closings by reducing net revenues. This is especially true for a cyclical industry like steel.

Other factors are also important, however. Since the firm will not replace the aging capital with new equipment, one determinant of a plant's net revenues is the amount of maintenance the capital in place requires in order to operate (in other words, its durability). The firm will continue to bear maintenance expenditures as long as the capital generates enough revenue to cover both the additional expense and other variable costs. Obviously, the larger the maintenance expenditures, the more they reduce net revenues, and the less likely they will be worth making.[8]

A low salvage value may also delay a plant's closing. The salvage value is the net amount of money the firm will realize when the plant closes. A large positive value means that much of the capital can be extracted without loss from the plant, thus shortening the time to shutdown. A negative value extends the time before exit, causing the plant to be operated even though total variable costs are not covered. In this situation, the firm would actually borrow to pay the uncovered variable costs in order to avoid the greater loss of closing.[9]

In general, the salvage value is determined by a plant's resale value minus costs incurred during closing. The resale value of the capital depends on its specificity to the production process and on output growth in the industry. The closing costs include the resources necessary to gather the information to make the closing decision and the time spent planning and executing it. The firm may also face employee-related closing expenses, such as severance pay, early retirement pay, and pensions, depending on previous contractual agreements or on local plant-closing legislation. Increases in these costs, by raising closing costs, will delay shutdowns.[10]

Thus, in a contracting industry with durable and specific capital and high closing costs, firms will delay closing plants. The plants exit eventually, but only after a long period of disinvestment. The result of selective and drawn-out disinvestment is a gradual increase in the average age of the industry's capital stock and a slowing of productivity growth.

Two things are vital to remember, however. First, in an industry with high exit barriers, a slow decline is the optimal rate of closure, despite years of poor earnings by the industry. Resources are always being utilized in their highest return activity during a contraction. Second, although an industry may appear to be failing because of lack of reinvestment, the antiquated plants are the result of exit barriers' prolonging exit and are not the cause of the industry's decline. While some plants will be modernized, those that are exiting will receive little investment.

In sum, an important consequence of allowing the market to reallocate resources from an industry with high exit barriers is that capacity will contract slowly, with old capacity lingering on and plants closing in bunches during downturns that lower revenues.

■ **7** In this context, fixed costs refer to costs that must be paid whether the plant is open or closed.

■ **8** See Lamfalussy (1961) for a discussion of these issues.

■ **9** The cost of going bankrupt, instead of continuing to pay uncovered variable costs, would be an upward bound on the amount the firm would be willing to borrow in this situation.

■ **10** This conclusion depends on the simplifying assumption made here that closing costs do not increase over time. As pointed out by Littman and Lee (1983), if employee-related closing costs rise quickly with the seniority of the work force, then a firm might accelerate closing to avoid the greater future liability.

III. The Size of Exit Barriers in the Steel Industry

Clearly, the magnitude of exit barriers in an industry depends on three factors: how long gross revenues are expected to cover variable costs, how specific and durable the capital is, and how high closing costs are.[11] This section presents some information about these factors in the steel industry which suggests that exit barriers are large.

TABLE 1

	1976	1986
Total Variable Cost of Raw Steel Materials, Energy and Labor (per net ton)	$217.00	$206.00
Total Variable Cost of Finished Steel Materials, Energy, and Labor (per net ton of finished product)	310.28	348.00
Total Cost of Finished Steel (per net ton of finished product)	361.38[a]	449.00

a. The number cited here is slightly lower than the figure reported by the Council on Wage and Price Stability, but is calculated as they describe in the text.

SOURCES: U.S. Council on Wage and Price Stability (1977), p. 60; Wharton Econometrics (1987), p. 4.5.

A rough idea of the likelihood that gross revenues will cover variable costs—the costs of all variable inputs to production—can be obtained by comparing the average variable cost of a ton of steel to the prices of various steel products. This cost is conventionally measured as the sum of labor, energy, and materials. The U.S. Council on Wage and Price Stability calculated that the average total variable cost per net finished ton of steel in 1976 was $310.28. Wharton Econometrics estimated that this cost equaled $348.00 in 1986. These estimates include the cost of producing raw steel, as well as the average industry cost of finishing it. Both of these studies also include estimates of the financing costs of steel production, taken here to be the average fixed cost of production (see table 1).

Table 2 compares estimates of average variable cost and average total cost for selected steel products to the average realized price per net ton of those products in 1976 and in 1986. In most cases, product prices were above the average variable cost. On the other hand, almost all of these prices were well below the total cost of finished steel. (Product prices do vary cyclically, causing the size of this shortfall to change over time. See table 3.) Overall, the data indicate that product prices may fall considerably below the average total cost without making immediate shutdown a firm's loss-minimizing alternative.

How long does a plant that is not covering total cost continue to operate? As stated above, unless prices dip or variable costs rise unexpectedly, a plant's closing would depend on the durability of its capital, on its resale value, and on the amount of closing costs.

Of these three, the high cost of closing appears to be the most important exit barrier currently in the steel industry. When closing a plant, a firm records a charge for the costs of dismantling the mill, for the operating loss until closing, for losses involved with contract terminations, and for a write-down of the assets. It also records the estimated liability for current and future payments to employees for pensions and insurance benefits.

The payments due to the work force when an integrated steel plant closes are substantial. For instance, by the provisions of a typical labor contract, qualified union members on layoff because of a permanent closing are eligible for severance pay, supplemental unemployment benefits, pension payments and, in some cases, supplemental pension payments.[12] Severance pay for union members with at least three years of seniority equals four to eight weeks' wages, depending on their years of service. A firm continues to pay life- and medical-insurance premiums for six to 12 months for workers with at least two years of continuous service. Workers may also be entitled to supplemental unemployment payments for up to two years.

One of the largest parts of the employee-related closing costs is the estimated liability for future pension payments. Of course, the portion of closing costs represented by the pension liability is not *caused* by closing, since the firm

■ **11** See Caves and Porter (1976) and Porter (1976) for an exhaustive list of various possible exit barriers. The types of barriers discussed here are those that seem particularly pertinent to the steel industry.

■ **12** The contract described here became effective in 1980. Terms of contracts made in later years appear to be quite similar.

owes retiring workers their pensions if the plant stays open. Nor are all of these charges out-of-pocket expenses. But they do represent payments that the firm must fund from some new source, since the cash flow from the plant will cease. This places an increased burden on a firm's remaining mills.[13]

TABLE 2

1976

	Average Variable Cost	Average Realized Price	Average Total Cost
Hot-Rolled Sheets	$282.30	$229.43	$333.40
Cold-Rolled Sheets	328.94	288.43	380.04
Hot-Dipped, Galvanized Sheets and Strip	356.92	368.59	408.02
Hot-Rolled Bars	286.96	311.14	338.06
Structurals	272.97	358.94	324.07

1986

	Average Variable Cost	Average Realized Price	Average Total Cost
Hot-Rolled Sheets	$305.00	$273.04	$406.00
Cold-Rolled Sheets	376.00	418.21	477.00
Hot-Dipped, Galvanized Sheets and Strip	419.00	537.93	520.00
Hot-Rolled Bars	313.00	360.03	414.00
Structurals	291.00	321.57	392.00

Note: The cost data from table 1 were adjusted for variation in finishing costs among products using data from Wharton Econometrics (1987), p. 4.7. Estimates are industry averages; costs are bound to be higher in exiting plants.

SOURCE: Bureau of the Census, *Current Industrial Reports: Steel Mill Products*, various issues.

In addition, because of the terms of pension agreements in this industry, the pension payments are actually higher if workers retire from a closing plant rather than from an operating mill. Under normal circumstances, union members are eligible for pensions after 30 years of service, or at age 65 (with 10 years of service), or at age 60 (with 15 years of service). But for workers laid off by plant closings, the eligibility requirements are eased. For instance, workers over 55, whose age plus years of service equal at least 70, become eligible. Also, some workers receive supplemental pension payments of $400 per month until they reach age 62, if they are laid off by a shutdown.

By the terms of this typical labor contract, it is clear that the size of the payments depends crucially on the age of workers and on their years of service. A firm might be able to reduce the work force somewhat by attrition before closing a plant, but under a seniority system, the remaining workers would tend to be older, with more years of service, which would drive up closing costs.[14]

These claims raise the cost of closing steel facilities enormously. In 1979, the United States Steel Corporation shut down a variety of mills and parts of mills, laying off more than 11,000 workers. According to the company's annual reports, the total cost of the closings was approximately $650 million, of which about $415 million represented labor-related expenses, implying a cost per worker of more than $37,000. Bethlehem Steel reported similar figures in its annual report, recording a $700 million liability in 1982 when about 18,000 workers were laid off during a restructuring that dealt principally with steel facilities.

More recent estimates show that these costs may be higher. One study indicates that the total cost per employee of closing a mill is $75,000, of which $54,000 represents employee-related closing costs (Wharton Econometrics [1987]). Using these figures, the Bethlehem Steel restructuring would currently cost $1.35 billion.

Firms cannot depend on high resale values to cover the large closing costs. The capital is quite specific to the industry and is of little value for any purpose other than steelmaking. Nor are other steel firms particularly interested in buying these plants; most integrated firms are reducing their capacity, and minimills are building new plants. Furthermore, the equipment in a closed plant is usually in need of major investment, since the former owner has disinvested from it before closing.[15]

■ **13** The problem is similar to that of Social Security when future generations are smaller. While in 1977 there were 2.3 workers for each retiree, currently there are two retirees for every steelworker.

■ **14** It is difficult to evaluate how these employee-related costs change over time. The severance payment formula does not appear highly sensitive to the seniority profile of the plant: the maximum severance payment is earned by workers with 10 years of experience. The supplemental pension payment is more complicated. The liability would increase if the number of qualifying workers rose over time (workers qualify if their combined age and years of service is over a certain minimum), and would fall if the number of qualifying workers fell over time (workers receive the payment only until age 62).

TABLE 3

Year	Hot-Rolled Sheets	Cold-Rolled Sheets	Hot-Dipped Galvanized Sheets & Strip	Hot-Rolled Bars	Structurals
1976	$229.43	$288.43	$368.59	$311.14	$284.46
1977	254.15	320.51	392.72	337.23	293.41
1978	281.10	354.31	430.35	364.26	326.98
1979	314.87	388.78	468.76	403.38	372.02
1980	317.30	395.42	487.64	415.90	408.03
1981	350.12	436.77	532.31	445.83	428.82
1982	338.79	433.87	525.84	414.94	421.90
1983	325.53	437.93	525.87	387.38	362.64
1984	326.01	453.18	560.16	393.49	358.52
1985	310.35	437.97	536.75	366.89	332.57
1986	273.04	418.21	537.93	360.03	321.57

SOURCE: Bureau of the Census, *Current Industrial Reports: Steel Mill Products,* various issues.

During the industry's contraction, there have been few examples of closing plants sold for continued operation as integrated steel mills. (One notable exception is the plant in Weirton, West Virginia. The employees purchased this mill from National Steel and have continued integrated production.) When sales do take place, the purchasers are generally interested in the rolling and finishing facilities, and keep steel furnaces closed. For instance, California Steel now imports semifinished steel for finishing at a (formerly integrated) plant in Fontana, which it purchased from Kaiser Steel.[16]

There are few opportunities to sell individual pieces of equipment. One company reportedly auctioned off some equipment when it went bankrupt, and some used equipment has been sold abroad, but no steelmaking operations have been sold for movement. Inventories of raw materials and parts can be distributed to other plants, but beyond that, the equipment is likely to sit until the price of steel scrap rises enough to pay the junk dealer for dismantling it.

The last exit barrier, the durability of steel industry capital, also works to delay plant closings by allowing the continued operation of aging equipment without major reinvestment. Furnaces and mills are depreciated over 15 to 20 years, but may operate for longer. For example, table 4 indicates that the *average* ages of various pieces of capital were more than 10 years in 1979, and that a significant percentage of the equipment had been operated for more than 20 years.

Of course, operation of the equipment still involves noncapitalized maintenance and repair expenditures. In addition, the blast furnaces, which provide the flow of hot metal to the steel furnaces in an integrated plant, require periodic relining. Blast furnaces basically operate on a continual basis for two to eight years, depending on their rate of utilization. But eventually the refractory material that prevents the hot metal from destroying the furnace must be replaced. Figures cited for a somewhat short-term repair process, called gunning, range from $14 million to $18 million. Actual replacement of refractories may cost anywhere from $20 million to $100 million, depending on the extent of the replacement and furnace rebuilding, though on average the cost will probably fall into the $20 million to $50 million range.

Frequently, firms will postpone a reline and leave the blast furnace idle, provided they have another operating furnace. But there are some limits to their ability to escape both operating losses and closing costs by idling entire plants. For instance, after being laid off for two years because of idled equipment, workers eligible for

■ **15** From 1960 to 1981, the average annual investment per ton of capacity in major pieces of steelmaking equipment was $34.08 in plants whose closing was announced before 1984, compared with $128.27 for plants remaining open (Deily [1988]). See Deily also for evidence that steel firms channeled investment away from plants that were least able to compete with imports and minimills, particularly during the period 1971-1981.

■ **16** See Wharton Econometrics (1987), p. 1.8, and J. Ernest Beazley, "Big Steel's Push to Extend Import Quotas Draws Debate," *Wall Street Journal,* December 30, 1987.

pensions may claim them. Also, laid-off workers are eligible for supplemental unemployment benefits for up to two years.

In sum, integrated steel firms appear to face sizable exit barriers. High closing costs, consisting principally of payments to employees, currently appear to be the largest barrier. Durable capital and low resale values also work to delay plant closings.

TABLE 4

	Average Age of Capacity (years)	Capacity Over 20 Years Old (percent)
Coke Ovens	17.3	46.9
Basic Oxygen Furnaces	11.0	2.3
Electric Furnaces	14.3	25.3
Hot Strip Mills	19.0	31.5
Aggregate[b]	17.5	33.3

a. As of January 1, 1979.

b. Includes data on open hearth furnaces, plate mills, wire rod mills, cold strip mills, and galvanizing lines.

SOURCES: American Iron and Steel Institute (1980), p. 21. Based on data from *The World Steel Industry Data Handbook,* vol. 1, and the American Iron and Steel Institute.

IV. Implications for Public Policy

The data presented here suggest that the decline of the steel industry has been painful and prolonged because of large closing costs and high exit barriers created by the technology of the production process. Although these barriers have delayed closings, resulting in lower profits and antiquated capital stocks in some plants, the necessary reduction of U.S. integrated steel capacity has been taking place, albeit slowly.

Is there any need for policies aimed at raising or lowering exit barriers? Although different firms, workers, stockholders, and communities could gain or lose, it is not at all clear that the economy as a whole benefits from either hastening or delaying plant closings. However, public policy in at least two areas of recent concern may have a strong impact on the steel industry's exit barriers.

First, the pension-insurance program affects exit barriers in the steel industry by altering the cost of closing plants. The Pension Benefit Guaranty Corporation (PBGC), a federally chartered agency that insures all workers with defined-benefit pensions, has already assumed some of the industry's plant closing costs and may ultimately assume more. As stated previously, pension liabilities are a major part of the cost of closing. A firm that desires to close plants, but that cannot afford to do so, may find that declaring bankruptcy is the cheapest way to reduce capacity, because the PBGC becomes responsible for the firm's pension liabilities.[17]

Thus, at least potentially, the PBGC could end up paying the pension liability portion of some firms' closing costs, thereby speeding up plant closings by lowering this particular exit barrier.[18] The situation has become more uncertain, however, because of the recent and still-unresolved differences between the PBGC and LTV Steel over responsibility for the latter's pension liabilities. Since this uncertainty makes it more difficult for firms to evaluate plant-closing decisions, it is important for policymakers to clarify who will ultimately pay these liabilities.

Policies to protect the industry from imports, on the other hand, may raise the exit barriers that steel firms face. The industry is currently protected by five-year Voluntary Restraint Agreements that the Reagan administration has negotiated with a number of steel-exporting countries. In the short run, the effect of the quotas may be to delay plant closings if the protection causes the industry to upwardly revise the expected revenues of its plants.

The long-run effects of the legislation are less clear. Firms are unlikely to reverse their long-run disinvestment from marginal plants unless they are convinced that the profitability of these plants has increased permanently. Such an assurance would require at least that the government make a long-term commitment to trade protection for the industry. But such a commitment would be expensive for domestic industries that use steel, and would by no means rule out further capacity reductions, since the minimill sector will continue to grow.

■ **17** See Buynak (1987) for a description of the limits on the amount of the firm's assets that the PBGC can claim to cover unfunded pension liabilities.

■ **18** Indeed, since the maximum payment the PBGC makes to workers may be well below workers' contracted pensions, and since supplemental payments for early retirement are not covered, the total cost of closing plants would be lower, though at the direct expense of the employees.

What public policy should not be doing is forcing reinvestment in the steel industry. The most misguided aspect of the trade protection currently in place is its requirement that the industry reinvest its net cash flow from steel businesses back into steel plants (Steel Import Stabilization Act of 1984, 19 U.S.C. 2253). The result of this directive may be to force investment in plants that will never yield an adequate return, a circumstance that will increase plant owners' losses when the plants are eventually closed.

Lack of reinvestment is not the underlying problem of the steel industry. Although investment in the plants that will survive is essential to their competitiveness, it is clear that additional capacity will eventually close. But shutdowns will be delayed as long as steel firms find that exit barriers make continued operation of marginal plants less costly than closing.

REFERENCES

Barnett, Donald F., and Robert W. Crandall. *Up from the Ashes: the Rise of the Steel Minimill in the United States.* Washington, D.C.: The Brookings Institution, 1986.

Barnett, Donald F., and Louis Schorsch. *Steel: Upheaval in a Basic Industry.* Cambridge, MA: Ballinger Publishing Co., 1983.

Beazley, J. Ernest. "Agency in Crisis: Bankruptcy Filings in Steel Overwhelm U.S. Pension Insurer." *Wall Street Journal.* May 21, 1987, p. 1.

Buynak, Thomas M. "Is the U.S. Pension-Insurance System Going Broke?" *Economic Commentary.* Federal Reserve Bank of Cleveland. January 15, 1987.

Caves, Richard E., and Michael E. Porter. "Barriers to Exit." In R.T. Masson and P. David Qualls, eds., *Essays on Industrial Organization in Honor of Joe S. Bain.* Cambridge, MA: Ballinger Publishing Co., 1976.

Crandall, Robert W. *The U.S. Steel Industry in Recurrent Crisis: Policy Options in a Competitive World.* Washington, D.C.: The Brookings Institution, 1981.

Deily, Mary E. "Investment Activity and the Exit Decision." *Review of Economics and Statistics.* (1988—forthcoming).

Kawahito, Kiyoshi. *The Japanese Steel Industry, with an Analysis of the U.S. Steel Import Problem.* New York: Praeger, 1972.

Lamfalussy, Alexandre. *Investment and Growth in Mature Economies: The Case of Belgium.* New York: St. Martin's Press, 1961.

Lawrence, Robert Z., and Robert E. Litan. *Saving Free Trade: A Pragmatic Approach.* Washington, D.C.: The Brookings Institution, 1986.

Littman, Daniel A., and Myung-Hoon Lee. "Plant Closings and Worker Dislocation." *Economic Review.* Federal Reserve Bank of Cleveland. (Fall 1983): 2-18.

Miller, Jack R. "Steel Minimills." *Scientific American.* 250 (1984:5): 32-39.

Porter, Michael E. "Please Note Location of Nearest Exit: Exit Barriers and Planning." *California Management Review.* 19 (1976:2): 21-33.

Stigler, George J. *The Theory of Price,* Third Edition. New York: Macmillan Co., 1966.

United States Council on Wage and Price Stability. *Report to the President on Prices and Costs in the United States Steel Industry.* Washington, D.C.: U.S. Government Printing Office, October 1977.

United States Department of Commerce, Bureau of the Census. *Current Industrial Reports: Steel Mill Products.* Series MA-33B. Washington, D.C.: U.S. Government Printing Office, various issues.

United States Steel Corporation. *Pension Agreement Between United States Steel Corporation and United Steelworkers of America.* Effective July 31, 1980. Document supplied by the United Steelworkers.

Wharton Econometrics. *Restructuring and Revival: The World Steel Industry, 1987-2000.* Volume III, Part II. Bala Cynwyd, Pennsylvania: Wharton Econometrics, 1987.

Why Do Wages Vary Among Employers?

by Erica L. Groshen

Erica L. Groshen is an economist at the Federal Reserve Bank of Cleveland. The author would like to thank John T. Dunlop, Richard B. Freeman, and Lawrence H. Summers for their comments and support on earlier versions. The paper also benefited from suggestions by Mary Deily, Randall W. Eberts, and Lawrence F. Katz.

Introduction

In neoclassical economics, wage rates—like the price of any traded commodity—are determined by both supply and demand. Despite the simultaneous nature of the wage-setting process, recent empirical investigations of the determinants of wages have focused primarily on factors affecting labor supply. Demand factors have been relatively neglected.

During the 1940s and 1950s, participation in the administration of wage and price controls led a distinguished group of economists to examine employer wage policies. Reynolds, Segal, Dunlop, Myers, Lester, and Lewis studied interindustry, intra-industry, union, establishment size, and regional differentials.[1] In essence, they focused on variables controlled by employers (that is, *labor demand*) and medium-run labor supply. Dunlop (1957) summarizes many of these effects in his work on wage contours.

Research on the influence of supply-side factors was stimulated by the development of human capital theory (Becker [1964] and Mincer [1974]), and by the availability of household surveys, which gather more information on workers than on their employers. Since the 1960s, labor economists have primarily studied variables controlled by employees (that is, long-run *labor supply* factors) such as age, education, and experience.

In the Current Population Survey, a household survey, regressions of wages on workers' characteristics typically produce results similar to those shown in table 1. In this example, the explanatory power of human capital variables is enhanced by exclusion of agricultural workers and of the youngest and oldest workers from the sample. Even within this limited population, the narrowly defined human capital variables explain only a quarter of the variation in the log of wages.[2] Addition of occupation raises explanatory power by 16 percent, while race, sex, and union variables add another 6 percent. Industry (broadly defined) raises explanatory power to 51 percent of the variation of wages.

What accounts for the 49 percent of wage variation that the equation doesn't explain? Are there other empirical regularities or theories that

■ **1** Segal (1986) and Kerr (1983) summarize the work of these economists.

■ **2** The explanatory power of human capital variables reported in table 1 is actually relatively high compared to that found in many samples because of exclusion of younger and older workers and of agricultural workers.

TABLE 2

Authors and Year	Data	Relevant Conclusions
7. Wachtel and Betsey (1972)	Survey of Consumer Finances (1967), Institute for Social Research sample of full-time, full-year service and production workers	Residuals of human capital wage regressions (with age, sex, race, job tenure, education, and marital status) are highly correlated with industry-occupation, union status, city size, and region dummies. Conclude that these structural (demand-side) variables, especially industry-occupation, are important determinants of wages because of rigidities in the labor market.
8. Dalton and Ford (1977)	1970 U.S. Census sample	Industry earnings increase with concentration up to a ratio of 0.5, after which they are stable. Sex and race differentials are large and significant for high concentration industries, while industry growth rate affects wages only in the more competitive industries. Regional differentials were significant but had changed since 1960.
9. Pugel (1980)	IRS profits by 3-digit industry, merged with industry average demographic and market data	Workers receive 7 percent to 14 percent of total excess profits: some of which buys higher skills, the rest of which is rent.
10. Krueger and Summers (1986a,b)	CPS, May 1974, 1979 and 1984; Quality of Employment Survey 1977	Industry wage differentials do not disappear when controlling for measured or unmeasured differences in human capital or for compensating differentials. Consistent with efficiency-wage models, lower turnover and better performance are apparently characteristic of high-wage industries.
11. Dickens and Katz (1986, 1987)	Current Population Surveys — all nonunion respondents for 1983	Divided workers into 12 occupational categories, calculated industry wage differentials in raw data, fixed effects equations (with human capital) and from residuals of human capital equations. Found that industry differentials are large, persistent, and correlated across occupations and countries. They are also correlated with industry characteristics: percent male, average education, quit rates, and measures of product market power and profitability. Conclude that simple competitive models are not consistent with observed patterns.

While evidence on the source(s) of the differentials remains inconclusive, a strong link between industry differentials and industrial concentration (or profit rates) is found in all studies that search for it (Slichter, Garbarino, Reynolds and Taft, Dalton and Ford, Pugel, and Dickens and Katz), except Weiss. Krueger and Summers find links between differentials and the predictions of efficiency wage models (lower turnover and higher effort).

■ **4** A further example of the complexity of the subject is that this discussion assumes that most establishments operate within a single industry and their wages reflect the patterns of the industry alone. This is a simplification that abstracts from very real examples. For instance, drug shelf stockers in supermarkets are paid the low wages common to drug stores rather than higher supermarket rates. In these cases, even the establishment is too high a level of aggregation.

B. Within-Industry Differentials

Table 3 summarizes a selection of the empirical literature that provides evidence of the existence of large wage differentials among firms and among plants.[5] The first studies are case studies, where many of the issues explored singly below are investigated for a single labor market. The first two studies are particularly valuable because they use data with unusually rich information on both worker and firm characteristics. Both studies find significant differentials among firms. Reynolds concludes that firms select the general wage level on which they operate until forced to change. Rees and Schultz estimate the individual and establishment effect on wages for four groups of occupations and find systematic differences among firms that are not consistent across all occupations.

Mackay, et al., Nolan and Brown, and Brown, et al. are fairly recent case studies of English and Australian labor markets. They find that wage variation by plant is a large and fundamental component of wage dispersion, and that employer wage differences are persistent over time and are linked to plant performance.

Like the English and Australian studies, Groshen (1988a) focuses on the entire employer differential within industry rather than on the portion associated with a particular characteristic. She finds that a random switch in employer, within detailed occupational category and industry, is associated with an expected wage change of 12 percent. She also finds that employer size, gender composition, and industry sector are associated with wage level. However, it is unlikely that measures of human capital such as experience, tenure, or education explain the observed establishment differentials. Groshen (1988b) finds that these interemployer wage differences are virtually stationary over six years and present within a single metropolitan statistical area. Hodson matches U.S. household survey data with employer information and finds employer characteristics to be strongly significant predictors of wages.

Investigations of employer size and gender composition wage differentials, such as those listed in table 3, are a dimension of the work on employer differentials because they select one aspect of establishment differentials for examination. The explanations for these phenomena must also come from the theories explored below. The worker-quality differential studies, by Evans and Conant, are of interest because they argue against sorting by ability or human capital.

Finally, the last two intra-establishment studies suggest that although interoccupational differentials are compressed within establishments, they do have the same patterns. Thus, establishment effects are fairly, but not exactly, uniform across occupations.

In summary, these studies provide strong evidence that within-occupation interemployer differentials exist, and that they are associated with measurable attributes of employers, such as firm or plant size.

II. Sources of Wage Differentials Among Employers

This section summarizes five models that explain why an employer might pay a wage premium to all of its employees rather than to particular individuals. These theories are based on the rigorous models of particular economic relationships that have been developed since the 1960s. Virtually all of the ideas in the following discussion can be found in the work of earlier economists, but were later formalized by, and are here referenced to, other authors.

A. The Role of Employers in the Basic Model of Wage Determination

The point of departure for the models of employer wage effects listed below is basic Marshallian supply and demand. I begin by noting that in a perfectly competitive labor market with costless contracting and information, and with identical workers and jobs, no differentials based on differences in labor demand would arise.

Market labor supply is a function of leisure preferences, population supply, and training costs. Market labor demand is the horizontal sum of all employers' demand curves, that is, the marginal revenue product of hours worked. Under perfect competition in capital and labor markets, equivalent workers at equivalent jobs earn the same wage. An employer whose wages stray from the market rate will be forced out of business by loss of employees (wages set too

■ **5** For a survey of the literature and the empirical problems associated with measuring a related issue, the relationship between compensation and firm performance, see Ehrenberg and Milkovich (1987).

TABLE 3

Authors and Year	Data	Relevant Conclusions
CASE STUDIES AND MORE GENERAL STUDIES OF INTEREMPLOYER DIFFERENTIALS		
1. Reynolds (1951)	Case study of an urban blue-collar labor market based on worker interviews and data published by other sources	Plant wage-level depends on industry, unusual efficiency of plant or management, secure monopoly or oligopoly control of product market, and history of relative wages. Most wage movements occur uniformly within clusters of firms. Plants operate within a range of feasible wage rates, but movement within the band is difficult.
2. Rees and Schultz (1970)	Personnel records from 75 Chicago establishments on 13 occupations, white- and blue-collar, skilled and unskilled; interviews with management personnel and workers	Industry differentials vary in size and sign across occupations, and are smaller for skilled workers. No positive relationship between establishment size and wages, within occupation, industry, location, and controlling for work characteristics. Location differentials are uniform across occupation.
3. Mackay, et al. (1971)	Mean earnings and quit rates by occupation from personnel records for blue-collar workers in 66 engineering plants in Birmingham and Glasgow from 1959 to 1966.	Within occupation, inter-plant coefficients of variation ranged from 16 percent to 23 percent and rank order correlations (from 1959 to 1966) were about 0.9, except for laborers. Wages were negatively correlated with quits, but unrelated to changes in plant size. Investigations of causes led to rejection of sorting by human capital, of random variations, and of working conditions. Concluded that efficiency wages for quit rates and profit-sharing were most likely sources.
4. Hodson (1983)	Wisconsin 1975 survey of high-school graduates from 1957, matched with employer information	Corporate structure variables (size, international links, capital intensity) strongly affect wages. Product market variables (profits, productivity) have little impact.
5. Nolan and Brown (1983)	10-year survey of wage structure for seven occupations in 25 factories in West Midlands, England	Employer effects on wage changes dominate occupation effects. Nevertheless, rankings by employer are relatively stable across occupation over 10 years; rank correlations of 0.8 to 0.9.
6. Brown, et al. (1984)	Survey of 44 occupations in 198 plants in Adelaide, Australia	Overawards to Australian workers tend to be tied to establishment rather than to occupation. Industrial concentration is highly correlated with size of overawards.
7. Groshen (1988a)	BLS Industry Wage Surveys of production workers' wages in six manufacturing industries	Within detailed job classification, wage variation between establishments accounts for 30-60 percent of wage variation, generating a standard deviation of 11 percent. Half of the differentials were associated with characteristics of the establishments (size, union affiliation, etc.).

TABLE 3

Authors and Year	Data	Relevant Conclusions
8. Groshen (1988b)	BLS Area Wage Surveys of nonsupervisory workers' wages (blue-collar and white-collar) in one SMSA for six years	Within detailed job classification, wage variation between establishments accounts for 20-70 percent of wage variation, generating a standard deviation of 12 percent. Differentials were unchanged over six years and not associated with growth or shrinkage.
WORKER QUALITY DIFFERENTIALS, WITHIN OCCUPATION, BETWEEN ESTABLISHMENTS		
1. Evans (1960)	Private area wage and salary survey of clerical workers in Boston	Across establishments, the strongest observed relationship was between wages and length of service. Test scores and education are inconsistent predictors of wages.
2. Conant (1963)	Placement test scores and beginning salaries for typists in Madison, WI	Test scores accounted for only 10 percent of the variation in starting wages offered by different employers to entry level typists.
ESTABLISHMENT AND FIRM SIZE DIFFERENTIALS		
1. Perlman (1940)	BLS Establishment Surveys—Wages and Hour Statistics for six industries	Hourly earnings are higher in large firms, within industry, occupation, product group, and region. Hourly earnings are not affected by establishment size, holding region constant.
2. Lester (1967)	BLS Industry Wage Survey and Census of Manufactures	Except for textiles, apparel and aircraft, earnings increase with establishment size. Differentials increase when fringe benefits are included.
3. Masters (1969)	BLS Census of Manufactures	Plant size variable is a stronger (larger and more significant) determinant of average wage differences among industries than concentration.
4. Buckley (1979)	BLS Area Wage Surveys for 29 areas	Controlling for industry mix, wages rise with area cost of living, but not with establishment size.
5. Miller (1981)	BLS Census of Manufactures	Controlling for industry, wages increase with size of establishment.
6. Personick and Barsky (1982)	BLS National Survey of Professional, Technical, and Clerical Pay 1980	Pay levels tend to increase with employer size, but above-average levels are associated only with large firms. Wage premia attributable to a firm's size are larger for entry-level than for experienced professional workers. Corporate size has better explanatory power for professionals while establishment size does better for clerical workers.
7. Mellow (1982)	Current Population Survey 1979	Both plant size and firm size are positively associated with wages, controlling for personal characteristics and concentration. The effect is proportionately larger when fringe benefits are included. Industry-plant size interaction variables were insignificant.

TABLE 3

Authors and Year	Data	Relevant Conclusions
8. Dunn (1980, 1984)	Independent surveys of employee wages, working conditions, and employer size within one industry	Large firms pay higher wages and shift premia than small firms, except in the highest-paid occupations. Compensating differentials do not appear to be the cause; infers the presence of bargaining.
9. Brown and Medoff (1987)	Variety of public sources	Firm and plant size are associated with higher wages, controlling for occupation, industry, and working conditions. Differentials are smaller for higher-grade occupations.
MALE/FEMALE COMPOSITION OF OCCUPATIONS WITHIN FIRMS		
1. Blau (1977)	BLS Area Wage Surveys	Within occupation, establishments tend to be segregated by sex; pay rates are negatively associated with percentage of establishment female. Occupational segregation by sex is associated with industry.
INTRA-ESTABLISHMENT OCCUPATIONAL DIFFERENTIALS		
1. Ward (1980)	BLS Area Wage Surveys	National occupational wage spreads do not exactly mirror individual firms; pay differentials are smaller within establishments.
2. Van Giezen (1982)	BLS Area and Industry Wage Surveys	Area occupational differentials are larger than intra-firm differentials. Intra-firm differentials vary by industry and region, and decrease with establishment size, although differences are small.

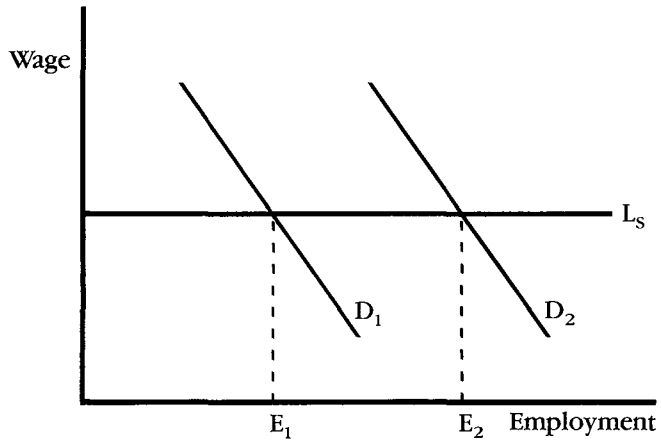
low) or the loss of capital (wages set too high).

The position that employers are price-takers is the theoretical basis for the current focus on labor supply as the only relevant determinant of wages. The employer in a competitive labor market faces a horizontal labor supply curve, as shown in figure 1. In the figure, Employer 1 has labor demand curve D_1, which differs from the labor demand curve of Employer 2 (labeled D_2). However, because they face a flat labor supply curve (L_S), the differences between the two employers affect only their employment levels (E_1 versus E_2), not their relative wages. Thus, the simple competitive model generates an empirically testable prediction: variations in labor demand should affect only quantity demanded, not wage level. This is true so long as demand differences do not affect worker utility.

The empirical work summarized above suggests that this simple model does not hold. Wages do vary among employers. In order to extend the simple model to allow for apparent demand-side effects, any explanation of wage variation by employer must answer two crucial questions: (1) why would one employer choose to pay more than another, and (2) why don't high-wage employers go out of business?

The answer to the first question is usually that a firm paying higher wages employs more productive workers. The advantage of the productivity explanation is that it also answers the second question. The disadvantage is that productivity differentials are usually due to individuals' abilities, not to employers' characteristics, implying the need for more explanation. If productivity differentials are not invoked, costly information or imperfect competition in the product market must be present and, again, operate similarly on all individuals in an establishment.

FIGURE 1



Source: Author.

B. Five Models of Employer Differentials

Table 4 summarizes five microeconomic sources of wage variation. Each source is developed from the competitive model by the introduction of transaction costs and/or of heterogeneity among agents. The table also lists the basic assumptions beyond those of the competitive model, and the additional assumptions necessary for the models to predict the existence of apparent *employer* wage differentials, rather than differentials among individuals or among occupations.

Each of the models examined predicts the existence of wage dispersion, and can be extended to predict employer-based dispersion, though the extensions usually involve extra twists of varying plausibility. Although none of the five models relaxes the assumption of profit maximization on the part of employers, they are arranged in order of their divergence from competitive theory in other aspects. In particular, the last two models, efficiency wages and bargaining, require assumptions of imperfections or lack of competition in the product or labor markets because they imply the existence of job rationing or queues for high-wage employers.

1. Sorting by Ability: Innate Differences, Human Capital, and Matching

The first two explanations relax the assumption of uniformity among workers or jobs in the market. Since the labor market is perfectly competitive, workers earn the marginal product of their work and employers pay equivalent wages per efficiency-unit of work. However, hourly wages may mismeasure either the workers' units of work (because this varies among workers) or their compensation (because it omits non-pecuniary returns to employment). In order to generate *establishment* differentials rather than just *individual* differentials, the theories must also explain why the marginal product of workers varies among employers.

Sorting models assume that some workers are more productive than others, and employers consistently hire their workers from a single quality stratum, regardless of occupation. The source of quality difference may be innate advantages (for example, genetic or motivational), or acquired differences (for example, education or work experience). Each establishment hires only the best, or only the worst, workers of each job category.

A priori, it is not obvious why an establishment would need or choose to segregate by ability. If all workers were paid their marginal products, the number of workers paid to produce a certain product should be irrelevant. For example, employers should be indifferent between two equally productive workers at one wage and a single doubly-productive worker at twice the wage. Any establishment could have a distribution of productivity levels (all rewarded accordingly) within each occupation. In this sort of world, no apparent establishment differentials would arise.

In order for innate or acquired productivity differences to generate apparent establishment differentials, employers must choose workers of fairly uniform productivity within occupations, and apply this policy similarly to all occupations. That is, this theory must be combined with an explanation for segregation by firm. Two questions arise: why and how?

The most convincing reason may be that employers' technologies are differentially sensitive to a worker's ability. In this case, employees of high ability who are not being rewarded for their higher ability by employers with ability-insensitive technology have an incentive to seek out employers who will pay according to ability. This leads to a positive correlation between the

ability-sensitivity of the employer's technology and the average quality of their applicant pool. Thus, employers with ability-sensitive technologies hire disproportionately more high-ability workers and, therefore, pay higher wages.[6]

For example, establishments requiring technical typing are likely to be highly sensitive to the skills of typists. So, we expect such employers to reward an excellent technical typist more than would employers who needed only text

TABLE 4

Model	Wage Equation[1]	Costly Factor(s)	Source(s) of Heterogeneity	Additional Assumptions Necessary for Existence of Employer Wage Effects
SORTING BY ABILITY				
Human Capital, Innate Differences, Job Matching	$w = MP$	Training	Innate or acquired worker quality, quality of job match	Establishments differ systematically by average quality of workers, or match, consistently across all or most occupations.
COMPENSATING DIFFERENTIALS				
Working Conditions, Fringe Benefits, Risk of Layoff	$w = MP$	Improvement of undesirable terms of employment	Management strategies or technologies	Undesirable terms of employment are uniform across all or most occupations within establishment.
RANDOM VARIATIONS				
Information, Search, Lagged Adjustment	$w = MP + \epsilon$ $\epsilon \sim f(0,\sigma^2)$	Employer and/or worker search, job mobility	Random draws from the pool, intertemporal wage variation	Employers vary in the average value of their draws, employers hire for all occupations during growth surges.
EFFICIENCY WAGES				
Monitoring, Turnover, Market Insulation, Corporate Consistency, Morale, Loyalty	$MP = f(w) \rightarrow w^* = MP^*$	Monitoring of workers' effort, turnover, design of internal wage structure, firm-specific training	Management strategies or technologies, corporate size	Employers adopt similar strategies (or technology has a similar effect) on the efficient wage across all or most occupations, workers in most occupations develop firm-specific training.
BARGAINING				
Insider/Outsider, Rent Capture, Gain-Sharing	$w = MP +$ $f(\pi$, workers' bargaining power)	Monitoring of workers and/or of management	Varying rents, ability of workers to capture rents, and/or managerial altruism	Rent capture is achieved and/or shared by all or most occupations.

■ **1** The symbols in this column are defined as:
w = wage
MP = marginal revenue product
ϵ = random error term, distributed with mean of *0* and variance of σ^2
f(*) = some function of *
w^*, MP^* = the unique profit-maximizing values of w and MP
π = profits

typing. The higher pay for skills will, in turn, attract other typists with technical skills into the applicant pools for employers needing technical typing. In order to create establishment differentials, this explanation must be expanded by the assumption that ability-sensitivity in establishments is highly correlated across occupations. Otherwise, wage variation would occur primarily by occupation within establishment, not by establishment across all occupations. Thus, in the example, the need for technical typing must be associated with ability-sensitivity in other occupations.

The second explanation is not mutually exclusive with the first and could provide a rationale for the correlation in ability-sensitivity across occupations. This model assumes that variation in the quality of workers in an establishment imposes negative externalities on the productivity of more able workers. Envision establishments as assembly lines where work stations are indivisible, or where the timing of the output depends on the speed of the slowest operative. Then, the productivity of the slowest worker determines the productivity of all the workers. As workers seek their best-paying job, establishments become segregated by quality.[7] Employers maximize profits by hiring or retaining (through their recruitment and termination policies) only those workers at least as able as those in their existing work force.

Job matching provides another approach within the sorting models (Jovanovic [1979]). Here, both worker and employer are uninformed about the worker's productivity in a particular job, until both have experienced it. The productivity of a worker-job combination is random, with a distribution known to both sides. Workers accept jobs that pay more than their current jobs. Employers offer wages based on the mean of the distribution, and later adjust wages to reflect measured productivity. Accuracy of productivity measurement improves as tenure increases. Employees with bad matches eventually leave in hope of finding a better match elsewhere. Then differences in the distribution of productivity across employers could lead to sorting.[8]

Other explanations for sorting come from the sociology literature on the joint productivity of teams as a product of the uniformity of team members. In all versions, all employers (whether high- or low-wage) earn zero or equal profits in equilibrium. But, high-wage/high-productivity employers are not associated with higher or lower profit levels than their low-wage/ low-productivity competitors. Only consistency matters.

The human capital model, formalized by Becker (1964) and Mincer (1974), provides a rationale for the variance of wages according to acquired training. Training increases productivity, raising the demand curve for hours of trained persons' time over that for untrained people. However, the costs of training, such as forgone wages and tuition, raise the supply curve for trained persons' time. Thus, the price of trained labor is higher than that of untrained labor and reflects the difference in marginal product between the two.

If human capital differences are manifested as employer differentials, employers must be able to predict productivity on the basis of acquired training (education and seniority), and both hire and pay workers accordingly. High-wage employers are such because they select the most highly trained workers in each occupational category. Low-wage employers hire (or end up with) workers with the least training across the board.[9]

Innate differences in productivity (for example, due to perseverance, or motivation) are less amenable to measurement by all parties, and are not included in the data bases generally available to economists. As such, they can only be investigated indirectly. However, if these innate qualities

■ **6** Models of self-selection and sectoral choice where the sectors vary in returns to ability in a competitive labor market were introduced in Roy (1951). A more recent treatment appears in Lang and Dickens (1987).

■ **7** When an employer pays wages that reflect actual marginal product, workers will be paid the marginal product of the least-productive worker, rather than according to their own individual abilities. Workers with higher potential will leave for jobs with a more productive "weakest link", causing average potential productivity to decline toward that of the least-productive worker. Employers who pay workers according to their potential marginal product will keep their workers, but lose money. This argument is similar to the "Jobs as Dam Sites" idea introduced in Akerlof (1981).

■ **8** For instance, suppose that all jobs had the same expected productivity, but those offered by certain employers had a higher variance. In this case, the high-variance employers might tend to have a high-wage, more-productive work force. This would happen because the workers with the good draws would stay longer and the workers with the worst draws would leave more quickly than they would in a firm with less variance.

■ **9** One explanation for sorting by establishment applies only to a particular form of acquired human capital: work experience. High-wage establishments may be older and have a relatively old, experienced work force, compared to the younger, less-productive workers in low-wage plants. If so, differences in age of employer would be reflected in wages, although wage per efficiency-unit of work is identical for all employers.

are correlated with the usual measures of acquired human capital such as age and experience, then controls for measures of human capital also control for innate differences.[10]

Conant (1963), Evans (1960), and Groshen (1988a) all suggest that employer wage differences are not associated with sorting by measured human capital or by ability correlated with human capital. Gibbons and Katz (1987) suggest that the unmeasured ability explanation also faces a number of empirical problems in addition to high correlation in employer differentials across occupations. One problem is the lower quit rates in high-wage firms and industries, which suggests that the high-wage jobs may be rationed, unless high ability has a particularly strong association with a tendency for employment stability. Another problem is that workers displaced from high-wage industries do not appear to retain their wage differentials if they switch industries. Finally, the correlation of employer wage differentials with product market power is difficult to explain within this model.

2. Compensating Differentials

The second possibility is compensating differentials, described by Adam Smith (1776), refined by other economists since then, and summarized in Smith (1979). The essential problem is mismeasurement of the total return to working. In the case of poor working conditions, monetary wage overstates the returns to individuals for their work because it ignores the extra costs imposed by working conditions.

Working conditions vary among employers, and it is costly to improve them. All else equal, workers prefer jobs with safe or pleasant working conditions to those with poor conditions. Thus, employers providing unfavorable conditions will be unable to meet their labor demand at the going wage. In response, the firms offering undesirable jobs must improve the working conditions or raise wages, whichever costs less. If improvement of conditions is costly, wages will be higher in order to attract sufficient labor, but the profitability of each hour worked is higher because of money saved during each hour worked under poor conditions.

If workers were identical, the wage differential between any two jobs would ensure that workers were indifferent between the two. If workers varied in their tastes, the differential would depend on the tastes of the marginal worker. The allocation of the work force among poor and good jobs depends on the assumptions made about existing production technologies. Technology is usually assumed exogenous, so we need a random distribution of differences in costs of improving conditions. If technology is not exogenous, all firms will choose the one that maximizes profits, so only those combinations of technologies and compensating differentials that yield the maximum profits will coexist.

In all versions of this model, employer (rather than individual) differentials arise only when quality of working conditions is consistent across all or most of the work force in establishments.[11] Many working conditions, such as physical exertion, do not apply because they are occupation-specific. However, high risk of layoff, poor ventilation, minimal fringe benefits, or inconvenient location could presumably affect all or most workers in an establishment. Then, the costs of improvement of these conditions must vary enough among employers to generate the large and persistent differentials.

Empirical studies of compensating differentials have been notably unsuccessful in finding evidence of their contribution to wage dispersion.[12] One exception to this generalization is Eberts and Stone (1985), who find evidence of compensating differentials only after controlling carefully for characteristics of employers, suggesting that compensating differentials are second-order effects. That is, type of employer determines overall level of compensation, but there is some substitution between wages and nonpecuniary compensation within groups of otherwise similar employers.

■ **10** Job market signalling (Spence [1973]) is an extreme example of this type of correlation, which blurs the distinction between human capital and innate differences.

■ **11** In addition, two fairly mechanical versions of compensating differentials are possible. The first is based on different age-earnings profiles with differing average tenure among plants. The second is variation in timing of annual salary adjustment. Groshen (1988a) presents evidence that suggests that neither of these possibilities is likely.

■ **12** For example, see Smith (1979). Most studies have attempted to identify compensating differentials among industries, where conditions vary most among employers. Nevertheless, such inquiries have been marked by their lack of success. For working conditions, see Brown (1980); for layoff risk, see Topel (1984). It is also unlikely that employer wage differences compensate for differences in fringe benefits. Freeman (1981), Smith and Ehrenberg (1981), and Atrostic (1983) find that inclusion of fringe benefits exaggerates wage differences among employers. That is, high-wage employers pay even more of total compensation in the form of fringe benefits than do low-wage employers.

3. Random Variations

Seminal articles by Stigler (1962) and Rothschild and Stiglitz (1976) launched a family of pure information models that use costly job search to explain wage dispersion. Suppose search were expensive for job-seekers. In this case the marketplace can sustain a range of wages because the gain from further search becomes uncertain, rather than a known quantity.[13]

In the typical model, establishments offer wages according to a distribution known to all job-seekers. Workers accept offers that exceed the expected value of further search. Job-rejecters pay to search again. Thus, the only sustainable distributions of wages are those where the minimum wage paid differs from the mean offer by less than the costs of employee search.

These models focus on the role of the individual in wage determination. No rationale is offered for variations among employers. A symmetric formulation of the problem from the employers' point of view posits the existence of a known distribution of reservation wages among a population of potential employees. Employers interview applicants to ascertain their reservation wages, and jobs are offered to workers (at their individual reservation wages) when the expected value of the wage reduction from an additional interview by the employer falls below the employer's search costs. Employer search costs consist mainly of advertising and interview expenses.

The employee-cost/employer-distribution model provides no theoretical basis for the existence of employer differentials. Rather, it explains only persistence of variance, leaving unanswered the question of why the employers who pay over the mean do not reduce their wages.

The converse model, the employer-cost/employee-distribution model, abstracts from the fact that firms usually set wages for a job rather than for an individual. Indeed, wages are usually attached to jobs before the interviewing process. Exceptions to this rule occur where job responsibilities are not well-defined, such as in very small firms and for highly skilled or very senior employees. In general, two individuals who differed only in reservation wage would not be offered different wages at the same plant.

Lagged adjustment, a second type of random variations model, is not inconsistent with the information/search models, but provides a basis for the variations (wage shocks) and an additional reason for their persistence (internal adjustment costs). These models, coined "geological models" by Dunlop (1982a), focus on the employer. Establishments may tend to hire in surges rather than in steady flows. If the costs of redesigning an internal wage structure are high or if workers are immobile, a firm's internal pattern and general level of wages will reflect the market wage pattern of its most recent expansion.[14]

In the random variation models, wages approximate the worker's marginal product, but costs of information introduce an error term. The mean of the error term is zero, and its variance is a positive function of the search and mobility costs for one or (perhaps) both parties. Consequently, establishment differentials result from random variations in the average error terms of employers. But, if establishment differentials are large, long-lived, and associated systematically with characteristics of employers—as suggested by the empirical work cited above—they are not random variations.

4. Efficiency Wages

Efficiency wage arguments posit a causal relationship between the wage level and a worker's on-the-job productivity.[15] Efficiency wage employers maximize profits by paying workers a premium above the market-clearing wage, because the resulting increment in productivity yields the highest profits. The increased productivity has been modeled as coming from three

■ **13** Originally, the information models were formulated to explain the existence of price or wage dispersion. Subsequent work uses these ideas to predict the level of unemployment. For example, see Azariadis (1983). Since the focus of the current work is wage dispersion, the earlier formulations of Stigler will be used to characterize the results of this diverse literature. Later versions of these models generate terminal wage distributions from initially assumed distributions. Stiglitz (1979) and Venables (1983) provide examples of these models.

■ **14** For example, establishments may grow by the addition of a second or third shift, rather than by hiring a few new workers each month. Wages at the time of a hiring surge reflect current labor-market conditions. If the market is tight, wages paid to attract new employees will be relatively high. Later, when market wages fall, adjustment down to the new market-clearing level will not be immediate. Redesigning the internal wage structure imposes costs (out-of-pocket and morale) on the employer. Wage schedules are rarely adjusted more often than annually and are rarely adjusted downward nominally. Upward adjustments will be slow if workers face mobility costs. Thus, the internal pattern and general level of wages at any particular time reflects the market wage pattern of the employer's most recent expansion. (Hence, the term "geological.")

■ **15** The main versions of these models are summarized in Yellen (1984) and Stiglitz (1984). Efficiency wages were originally formulated as an explanation for equilibrium unemployment, rather than for wage dispersion. Wages do not fall to clear the market because firms maximize profits in a labor market where wages are high and jobs are rationed.

sources: reduced monitoring (or shirking) costs, decreased turnover, and sociological considerations. The internal labor market literature adds two more possibilities: market insulation and corporate consistency.

In the monitoring/shirking version, workers' effort is costly to monitor (Bulow and Summers [1986], Shapiro and Stiglitz [1984]). An increase in wages decreases a worker's incentive to shirk, because shirking increases the probability of losing a high-wage job. In comparison to an employer paying the equilibrium wage, efficiency wage employers pay higher wages, experience higher worker productivity, and have lower direct monitoring expenses.

The turnover version emphasizes employer costs of hiring and training (Salop [1979]). Wages above equilibrium reduce turnover because workers have fewer superior alternatives and/or because the general level of unemployment rises. Thus, workers paid higher wages have longer tenure. Two related search/recruiting versions of the model show that firms with high costs of unfilled vacancies will offer high wages to more quickly fill vacancies (Lang [1987] and Montgomery [1987]).[16]

A third variant of the argument is based on sociological morale, loyalty, or teamwork effects. Group work norms are raised by wages above the minimum required. Akerlof (1982) terms this the "partial gift exchange" model.

The two internal labor-market variants, as described by Doeringer and Piore (1971), focus on the out-of-pocket and morale costs of designing a compensation package for a group of employees, and on firm-specific human capital. If all wages are to be set constantly at market-clearing levels, shocks to the external labor market will necessitate periodic readjustments of internal pay relationships. Yet, redesign of wage schedules may be expensive for certain types of employers, especially large ones, or for certain groups of employees, such as incentive workers. In addition, any change in relative wage relationships may be perceived as inequitable or as a breach of implicit contract. Such dissatisfaction could reduce productivity through increased shirking or turnover.

An alternative to frequent, disruptive adjustments in response to market fluctuations is to set wages above the market level. If, on average, workers receive a premium, then wage shocks that are small relative to the premium will not force a firm to readjust its compensation package. Employers save out-of-pocket and productivity costs of the adjustment, in return for paying higher wages.

Corporate consistency, the second internal labor-market version, is based on the tendency of large firms to promote workers from within whenever possible rather than hire from outside. Presumably, firm-specific human capital makes promotions or transfers among plants efficient. Nevertheless, such a policy requires that internal wages for each occupation in each plant meet two criteria: (1) they cannot be much lower than local wages for the occupation (or the workers will leave the firm), and (2) they cannot be lower than firm-wide wages for that occupation (or workers will refuse transfers to the plant). This implies identical wage structures for each plant within the firm regardless of location, as long as product lines are similar enough for personnel to be transferred among them. Furthermore, each occupation will earn the maximum local rate over all plant locations. On average, this yields positive establishment differentials that increase with firm size.

Efficiency-wage models can be invoked to explain differentials among firms in two ways. First, the profit-maximizing point is, almost by definition, locally flat. This implies the existence of a plane of (almost) iso-profit wage-productivity points for identical firms. That is, variations in wages from the optimum lead to only small profit losses. Firms are close to indifferent among the possible combinations, so a random distribution of strategies results (Bulow and Summers [1986]).

A second, more plausible, explanation stems from economically important heterogeneity among employers: differences in technology (vintage effects, for example), or differences in products (such as differentiated quality niches). The productivity of workers at the market-clearing wage may be indistinguishable from high-productivity work under some technologies, or may be adequate for one market but not for another. Workers paid the market-clearing wage form a queue for jobs at the elevated wage, while recipients of the high wage avoid job loss or job changes because of the scarcity of equivalent opportunities.

Efficiency differentials can explain establishment differentials when workers in all or most occupations in the establishment are affected. That is, it is crucial that the heterogeneity among employers affect the efficient wage for all

■ **16** Lang (1987) extends the analysis to show that an equilibrium distribution of wages can be sustained among otherwise-identical firms, but there is no reason to expect firms' positions in the distribution to persist, unless firms lock in their position by their choice of technology. This assumes the existence of a range of technologies, each with different capital-intensity (and, thus, cost of unfilled vacancy).

occupations similarly. The plausibility of this assumption depends on the version of the model in question.

Few empirical tests of efficiency wage models have been performed, primarily because of the lack of appropriate data. One recent exception, Leonard (1987), finds little evidence to support the turnover or supervisory-intensity versions among electronics companies in California. Another study, Krueger and Summers (1986a) finds some support for efficiency wage explanations of interindustry wage differentials. Interest in these models suggests that the results of other tests may be available shortly.

5. Insider/Outsider Bargaining Models

When bargaining between workers and their employers takes place in the context of competitive markets (in labor, capital, and products), bargained wages cannot differ from the market-clearing wage. Otherwise, the firm would close or the workers would leave. However, if employees can exercise a claim on the rents generated by an enterprise, they will bargain (implicitly or explicitly) with their employers. Wage settlements will reflect both the size of rents and the relative bargaining power of the parties. Thus, the existence of both rents to the firm and employee bargaining power are necessary conditions for wage bargaining to produce wage variation.

Although all versions of bargaining models must assume the existence of rents, the models differ in the identity of the bargaining agents and in the enforcement mechanisms for the bargaining. The bargaining agent for the workers is most clear in the case of unionism. In the collective bargaining literature, the outcome of negotiation is likened to the Edgeworth Box. Bargaining is a positive-sum game until the contract curve is reached, and a zero-sum game along the contract curve. The outcome is determined by the relative bargaining ability and credibility of participants' threats. The range of possible wages is bounded by the market-clearing wage on the bottom end and by the worker's actual marginal product (with labor appropriating all rents and capital earning the normal rate of return) on the high end.

In a nonunion setting, the bargaining agent for the workers is not obvious. However, economists have long noted the existence of informal organization by workers in nonunion settings (Dunlop [1957]). One version is the union-threat effect, where the threat of unionization forces owners to provide nonunion workers benefits similar to those they would receive if unionized (Dickens [1986]).

In a second version, the managerial-capitalism or agency-cost version, managers act as mediators between labor and the owners of capital. If the rewards to management are not highly correlated with rents to the owners, or if managers maximize a utility function dependent on worker satisfaction (whether due to managerial altruism or to the ability of workers to impose on-the-job problems), then management may not act to maximize rents to owners. Implicit bargaining may occur, with management cast in various roles from agent for the workers, to mediator between the two sets of interests, to agent for the owners. The latter role would generate a model all but institutionally indistinguishable from a union bargaining model. For example, Aoki (1984) presents cooperative bargaining models for modern nonunion corporate enterprises with various constituencies. Edwards (1979) also presents an informal model of nonunion bargaining.

Bargaining models easily lend themselves to the prediction of establishment differentials. The only additional assumption necessary is one that binds together workers of different occupations in the establishment. Three possibilities exist. First, workers' bargaining power may be consistent across occupations in an establishment. Second, perhaps workers must form large groups in order to exert bargaining power. Third, managerial altruism may extend uniformly across occupations.

The persistent link between measures of product-market power and industry wage differentials provides an empirical basis for further investigation of bargaining theories. More direct evidence is limited by the lack of data, but studies by Abowd (1985) on unionized firms and by Kleiner and Boullion (1987) on both union and nonunion firms provide some support for bargaining hypotheses.[17] As with efficiency wage models, more direct tests of these models are certain to be available in the near future.

■ **17** Abowd (1985) finds evidence that union contract settlements diminish the value of the firm by exactly the change in the value of the negotiated settlement. Kleiner and Boullion (1987) find that firms' wages are strongly positively correlated with the provision of sensitive financial information to employees.

III. Labor Market Policy and Employer Wage Effects

The empirical work cited in this paper suggests that employer wage differentials are large. Thus, they may account for many of the observed inequalities in the labor market, such as those among races or between men and women. Exploration of five models of employer differentials clarifies the point that these differentials are not necessarily inconsistent with profit maximization by firms acting in a competitive labor market. Yet each model suggests the existence of a particular barrier that prevents formation of a single market wage.

The link between theories of employer wage effects and labor market policy to reduce income inequality is labor-market segmentation.[18] When labor markets are segmented, workers are separated into distinct markets by institutional barriers that prevent workers or employers from switching between markets. Thus, different wages persist for each sector of the labor market. Although workers in each sector are paid their marginal product, productivity varies between sectors according to sector-specific supply and demand, or sector-specific quality. Obviously, the costs of barrier removal must be high enough to prevent profit-seeking employers from eroding the differences between sectors.

Employer differentials will create segmented markets only if employers limit their recruitment to one sector, so any model must explain why employers hire all (or most) of their employees from the same market sector. Each model discussed above introduces a barrier that could create segmentation, with strikingly different policy implications. Thus, it is precisely the identification of the source of the barrier that makes segmentation difficult to cure with policy.

For example, under the sorting model, segmentation will arise if workers of different sex or race have different access to human capital. The model implies a need for the development of human capital among secondary sector workers (for example, lower cost, better education, or job training). Alternatively, compensating differentials imply no role for policy, since the market actually remunerates all workers equally. Apparent segmentation arises simply because tastes differ systematically among groups.[19] Random variations suggests that search costs are higher for the classes of workers in predominantly low-wage jobs. A possible solution may be expansion of job-service agencies targeted to these groups.

Efficiency wages and bargaining imply the existence of queues of workers for high-wage jobs. Thus, any attempt to reduce inequality should rest on regulation of employers' recruitment policies, on improvement of placement services for secondary market workers, and on elimination of any minor productivity deficiencies among workers in the secondary sector.[20]

These five theories of wage determination also diverge from each other in their predictions for the impact of other kinds of policy. For example, Stiglitz (1984) and Bulow and Summers (1986) analyze the effects of efficiency wages on macroeconomic performance and trade policy. Weitzman (1986) offers an analysis of the effects of a particular form of profit-sharing on economic stability and growth.

Understanding the source of employer differentials is clearly important for understanding the distribution of wages, and for formulating policy to affect it. New sources of data must be developed to allow research on employer activities such as supervision, recruitment, terminations, and wage-setting. Without further research on these topics, we will remain unable to sort out whether employer wage differentials are signs of inefficiency, of discrimination, or of other market imperfections.

■ **18** For a summary of the literature on segmentation, see Cain (1976) and Dickens and Lang (1985). Lang and Dickens (1987) provide a detailed investigation of the relationship between the literature on segmented markets and neoclassical economic theory.

■ **19** For instance, compared to men, women may prefer quieter, cleaner, or more flexible jobs (Filer [1983]).

■ **20** Bulow and Summers (1986) demonstrate how efficiency wages may be a source of market segmentation. They emphasize that segmentation requires the existence of a small productivity differential between workers of the two sectors, but that the wage difference between the two sectors will be far greater than the productivity difference. A similar argument can be made for differentials associated with rent-sharing, assuming profit maximization on the part of employers.

REFERENCES

Abowd, John M. "Collective Bargaining and the Division of Value of the Enterprise." Cornell University. Unpublished paper, October 1985.

Akerlof, George A. "Gift Exchange and Efficiency Wage Theory: Four Views." *American Economic Review.* 74 (May 1984): 79-83.

_____. "Labor Markets as Partial Gift Exchange." *Quarterly Journal of Economics.* 97 (November 1982): 543-569.

_____. "Jobs As Dam Sites." *Review of Economic Studies.* 48 (January 1981).

Aoki, Masahiko. *The Cooperative Game Theory of the Firm.* New York: Oxford University Press, 1984.

Atrostic, B.K. "Alternative Pay Measures and Labor Market Differentials." U. S. Department of Labor Office of Research and Evaluation, Bureau of Labor Statistics Working Paper No. 127, March 1983.

Azariadis, Costas, "Employment With Asymmetric Information." *Quarterly Journal of Economics.* 98 Supplement (1983): 157-172.

Becker, Gary. *Human Capital.* New York: NBER, 1964.

Blau, Francine. *Equal Pay in the Office.* Lexington, Mass: D.C. Heath and Co., 1977.

Brown, Charles. "Equalizing Differences in the Labor Market." *Quarterly Journal of Economics.* 94 (February 1980): 113-134.

_____, and James L. Medoff. "The Employer Size Wage Effect." Unpublished paper, October 1987.

Brown, William, John Hayles, Barry Hughes, and Lyndon Rowe. "Production and Labor Markets in Wage Determination: Some Australian Evidence." *British Journal of Industrial Relations.* 22 (July 1984).

Buckley, John E. "Do Area Wages Reflect Area Living Costs?" *Monthly Labor Review.* 102 (November 1979): 24-29.

Bulow, Jeremy I., and Lawrence H. Summers. "A Theory of Dual Labor Markets With Application to Industrial Policy, Discrimination, and Keynesian Unemployment." *Journal of Labor Economics.* 4 (July 1986): 376-414.

Cain, Glenn G. "The Challenge of Segmented Labor Market Theories to Orthodox Theory." *Journal of Economic Literature.* 14 (December 1976): 1215-1257.

Conant, Eaton H. "Worker Efficiency and Wage Differentials in a Clerical Labor Market." *Industrial and Labor Relations Review.* 16 (April 1963): 428-433.

Dalton, James A., and E. J. Ford, Jr. "Concentration and Labor Earnings in Manufacturing and Utilities." *Industrial and Labor Relations Review.* 31 (October 1977): 45-60.

Dickens, William T. "Wages, Employment and the Threat of Collective Action by Workers." National Bureau of Economic Research Working Paper No. 1856, March 1986.

_____, and Lawrence F. Katz. "Industry Wage Differences and Theories of Wage Determination." National Bureau of Economic Research Working Paper No. 2271, July 1987.

_____, and Lawrence F. Katz. "Industry Wage Differences and Industry Characteristics." National Bureau of Economic Research Working Paper No. 2041, September 1986.

_____, and Kevin Lang. "Labor Market Segmentation and the Union Wage Premium." National Bureau of Economic Research Working Paper No. 1883, April 1986.

_____, and Kevin Lang. "A Test of Dual Labor Market Theory." *American Economic Review.* 75 (September 1985): 792-805.

_____, and Kevin Lang. "Testing Dual Labor Market Theory: A Reconsideration of the Evidence." National Bureau of Economic Research Working Paper No. 1670, July 1985.

Doeringer, Peter B. and Michael J. Piore. *Internal Labor Markets and Manpower Analysis.* Lexington, Mass: D.C. Heath and Co., 1971.

Dunlop, John T. "Fundamentals of Wages and Labor Markets." Unpublished paper, Harvard University. August 30, 1982(a).

_____. "What Both Sides Should Know About Compensation." Unpublished paper, Harvard University. August 30, 1982(b).

_____. "The Task of Contemporary Wage Theory." In G. Taylor and F. Pierson, eds., *New Concepts in Wage Determination.* New York, 1957.

Dunn, Lucia. "The Effects of Firm and Plant Size on Employee Well-Being." In John Siegfried, ed., *The Economics of Firm Size, Market Structure and Social Performance.* Washington, D.C.: Federal Trade Commission, 1980.

_____. "The Effects of Firm Size on Wages, Fringe Benefits, and Worker Disutility." In Harvey Goldschmidt, et al., eds., *The Impact of the Modern Corporation.* New York: Columbia Press, 1984.

Eberts, Randall W., and Joe A. Stone. "Wages, Fringe Benefits and Working Conditions: An Analysis of Compensating Differentials." *Southern Economic Journal.* 52 (July 1985): 274-280.

Edwards, Richard. *Contested Terrain.* New York: Basic Books, Inc., 1979.

Ehrenberg, Ronald G., and George T. Milkovich. "Compensation and Firm Performance." In Kleiner, Morris, et al., eds., *Human Resources and the Performance of the Firm.* Industrial Relations Research Association, Madison, Wis. (1987): 87-122.

Evans, Robert Jr. "Worker Quality and Wage Dispersion: An Analysis of a Clerical Labor Market in Boston." *Industrial Relations Research Association Proceedings.* 1960.

Filer, Randall K. "Sexual Differences in Earnings: The Role of Individual: Personalities and Tastes." *Journal of Human Resources.* 18 (Winter 1983): 82-99.

Freeman, Richard B. "The Effect of Unionism on Fringe Benefits." *Industrial and Labor Relations Review.* 34 (July 1981): 489-509.

Garbarino, Joseph W. "A Theory of Interindustry Wage Structure Variation." *Quarterly Journal of Economics.* 64 (May 1950): 282-305.

Gibbons, Robert, and Lawrence F. Katz. "Learning, Mobility, and Inter-Industry Wage Differentials." Massachusetts Institute of Technology. Unpublished paper, December 1987.

Groshen, Erica L. "Sources of Wage Dispersion: The Contribution of Interemployer Differentials Within Industry." Federal Reserve Bank of Cleveland Working Paper 8802, January 1988a.

_____. "The Size and Stability of Interemployer Wage Differentials Within an Area." Unpublished paper, 1988b.

Hodson, Randy. *Workers' Earnings and Corporate Economic Structure.* New York: Academic Press, 1983.

Holmstrom, Bengt. "Equilibrium Long-Term Labor Contracts." *Quarterly Journal of Economics.* 98 (Supplement, 1983): 23-54.

Jovanovic, Boyan. "Job Matching and the Theory of Turnover." *Journal of Political Economy.* 87 (October 1979): 972-990.

_____. "Firm-Specific Human Capital," *Journal of Political Economy.* 87 (December 1979): 1246-1260.

Kerr, Clark. "The Intellectual Role of Neorealists in Labor Economics." *Industrial Relations.* 22 (Spring 1983): 298-318.

Kleiner, Morris M., and Marvin L. Boullion. "Providing Business Information to Production Employees: Impacts on Compensation and Profitability." University of Minnesota. Unpublished paper, 1987.

Krueger, Alan, and Lawrence H. Summers. "Efficiency Wages and the Wage Structure." National Bureau of Economic Research Working Paper No. 1952, June 1986(a).

_____. "Reflections on the Inter-Industry Wage Structure." National Bureau of Economic Research Working Paper No. 1968, June 1986(b).

Lang, Kevin. "Persistent Wage Dispersion and Involuntary Unemployment." Boston University. Unpublished paper, December 1987.

_____, **and William T. Dickens.** "Neoclassical and Sociological Perspectives on Segmented Labor Markets." National Bureau of Economic Research Working Paper No. 2127, January 1987.

Leonard, Jonathan S. "Carrots and Sticks: Pay, Supervision and Turnover." *Journal of Labor Economics.* 5 (October 1987).

Lester, Richard A. "A Range Theory of Wage Differentials." *Industrial and Labor Relations Review.* 5 (July 1952).

_____. "Pay Differentials by Size of Establishment." *Industrial Relations.* 7 (October 1967): 57-67.

Lewis, L. Earl. "Wage Dispersion in Manufacturing Industries 1950-1955." *Monthly Labor Review.* (July 1956): 780-786.

Mackay, Donald I., David Boddy, John Brack, John A. Diack, and Norman Jones. *Labour Markets Under Different Employment Conditions.* London: George Allen & Unwin Ltd., 1971.

Masters, Stanley H. "An Interindustry Analysis of Wages and Plant Size." *Review of Economics and Statistics.* 51 (August 1969): 341-345.

Medoff, James L., and Katherine G. Abraham. "Experience, Performance and Earnings." *Quarterly Journal of Economics.* 95 (December 1980): 703-736.

Mellow, Wesley. "Employer Size and Wages." *Review of Economics and Statistics.* 64 (August 1982): 495-504.

Miller, Edward. "Variation of Wage Rates With Size of Establishment." *Economic Letters.* 8 (1981).

Mincer, Jacob. *Schooling, Experience, and Earnings.* New York: NBER, 1974.

Montgomery, James. "Equilibrium Wage Dispersion and Interindustry Wage Differentials." Massachusetts Institute of Technology. Unpublished paper, December 1987.

Nolan, Peter, and William Brown. "Competition and Work Place Wage Determination." *Oxford Bulletin of Economics and Statistics.* 45 (August 1983).

Perlman, Jacob. "Hourly Earnings of Employees in Large and Small Enterprises." Temporary National Economic Committee, United States Government Printing Office, Monograph No. 14, 1940.

Personick, Martin E., and Carl B. Barsky. "White Collar Pay Levels Linked to Corporate Work Force Size." *Monthly Labor Review.* 105 (May 1982): 23-28.

Pugel, Thomas A. "Profitability, Concentration and the Interindustry Variation in Wages." *Review of Economics and Statistics.* 62 (May 1980): 248-253.

Rees, Albert, and George P. Schultz. *Workers and Wages in An Urban Labor Market.* Chicago: The University of Chicago Press, 1970.

Reynolds, Lloyd G. *The Structure of Labor Markets.* New York: Harper & Brothers, 1951.

_____, **and Cynthia Taft.** *The Evolution of Wage Structure.* New Haven, Conn.: Yale Press, 1956.

Rosen, Sherwin. "On the Interindustry Wage and Hours Structure." *Journal of Political Economy.* 77 (March-April 1969): 249-273.

Rothschild, Michael, and Joseph E. Stiglitz. "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *Quarterly Journal of Economics.* 90 (November 1976): 629-649.

Roy, Andrew D. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers.* 3 (June 1951): 135-146.

Salop, Steven C. "A Model of the Natural Rate of Unemployment." *American Economic Review.* 69 (March 1979): 117-125.

Segal, Martin. "Post-Institutionalism in Labor Economics: The Forties and Fifties Revisited." *Industrial and Labor Relations Review.* 39 (April 1986): 388-403.

Seiler, Eric. "Piece Rate vs. Time Rate: The Effect of Incentives on Earnings." National Bureau of Economic Research Working Paper No. 879, April 1982.

Shapiro, Carl, and Joseph E. Stiglitz. "Equilibrium Unemployment as a Worker Discipline Device." *American Economic Review.* 74 (June 1984): 433-444.

Slichter, Sumner. "Notes on the Structure of Wages." *Review of Economics and Statistics.* 32 (February 1950).

Smith, Adam. *The Wealth of Nations* (1776). Penguin English Library Edition, 1982.

Smith, Robert S. "Compensating Wage Differentials and Public Policy: A Review." *Industrial and Labor Relations Review.* 32 (April 1979): 339-352.

Smith, Robert, and Ronald G. Ehrenberg. "Estimating Wage-Fringe Trade-Offs: Some Data Problems." National Bureau of Economic Research Working Paper, No. 827, December 1981.

Spence, A. Michael. "Job Market Signalling." *Quarterly Journal of Economics.* 87 (May 1973): 355-375.

Stigler, George J. "Information in the Labor Market." *Journal of Political Economy.* 70 (October 1962): S94-105.

Stiglitz, Joseph E. "Theories of Wage Rigidities." National Bureau of Economic Research Working Paper, 1984.

_____. "On Search and Equilibrium Price Distributions." In Michael J. Posken, ed., *Economics of Human Welfare.* New York: Academic Press, Inc., 1979.

Topel, Robert H. "Equilibrium Earnings, Turnover and Unemployment." *Journal of Labor Economics.* (October 1984): 500-522.

Van Giezen, Robert W. "A New Look At Occupational Wages Within Individual Establishments." *Monthly Labor Review.* 105 (November 1982): 22-28.

Venables, Anthony J. "Random Job Prospects and the Distribution of Income." *Quarterly Journal of Economics.* 98 (November 1983): 637-657.

Wachtel, Howard M., and Charles Betsey. "Employment at Low Wages." *Review of Economics and Statistics.* 54 (May 1972): 121-129.

Wachter, Michael L. "Cyclical Variation in the Interindustry Wage Structure." *American Economic Review.* 60 (March 1970): 75-84.

Ward, Virginia L. "Measuring Wage Relationships Among Selected Occupations." *Monthly Labor Review.* 103 (May 1980): 21-25.

Weiss, Leonard A. "Concentration and Labor Earnings." *American Economic Review.* 56 (March 1966): 96-117.

Weitzman, Martin. *The Share Economy: Conquering Stagflation.* Cambridge, Mass.: Harvard University Press, 1986.

Yellen, Janet L. "Efficiency Wage Models of Unemployment." *American Economic Review.* 74 (May 1984): 200-205.