EFFICIENCY AND FAIRNESS IN INSURANCE RISK CLASSIFICATION

*Kenneth S. Abraham**

THE heart of any insurance system is its method of classifying risks and setting prices. Different methods of classification can produce different safety incentives, different risk distributions, and different protection against loss. Because Americans spend over $250 billion annually on private insurance,[1] classification practices have enormous economic significance. These practices also have moral implications, because risk classification produces risk sharing. Yet the nature of risk classification and the economic and moral purposes it serves remain inadequately analyzed.[2]

The result has been a series of unresolved and, in the view of some,[3] unresolvable controversies concerning the risk classification

---

* Professor of Law, University of Virginia School of Law. This article is based on a chapter from my forthcoming hook, Distributing Risk: Insurance, Legal Theory and Public Policy (Yale University Press 1986). I am grateful to the Esther A. and Joseph Klingenstein Fund for research support. I also wish to thank the participants in workshops at the Universities of Maryland and Virginia and at Yale for their comments on an earlier draft. I am especially indehted to Guido Calabresi, Charles Goetz, Henry Hansmann, Jeffrey O'Connell, and Samuel Rea, Jr.

[1] In 1982, private health insurance premiums in the United States totaled almost $99 hillion. See Health Ins. Ass'n of Am., Source Book of Health Insurance Data 2 (1982-1983). In the same year, property/casualty insurance premiums totaled $104 billion. See Insurance Information Inst., Insurance Facts 6 (1983-1984). Life insurance premiums that year totalled approximately $50 hillion. See American Council of Life Ins., Life Insurance Fact Book 5 (1983).

[2] Scholarly studies have considered only specific forms of classification and have neither relied on nor attempted to elucidate principles of general applicability. The vast majority have focused on sex discrimination. See, e.g., Benston, The Economics of Gender Discrimination in Employee Fringe Benefits: *Manhart* Revisited, 49 U. Chi. L. Rev. 489 (1982); Bernstein & Williams, Title VII and the Problem of Sex Classifications in Pension Programs, 74 Colum. L. Rev. 1203 (1974); Brilmayer, Hekeler, Laycock & Sullivan, Sex Discrnination in Employer-Sponsored Insurance Plans: A Legal and Demographic Analysis, 47 U. Chi. L. Rev. 505 (1980); Kimball, Reverse Sex Discrimination: *Manhart*, 1979 Am. B. Found. Research J. 83; Rutherglen, Sexual Equality in Fringe-Benefit Plans, 65 Va. L. Rev. 199 (1979).

[3] Cf. Austin, The Insurance Classification Controversy, 131 U. Pa. L. Rev. 517 (1983) (insurance classification controversies cannot he resolved by current legal principles absent drastic changes in social stratification).

practices of the insurance industry. On the one hand, if two purchasers of insurance coverage have different predicted losses or actual loss experiences, it is easy to conclude that the prices that each pay should reflect those differences. On the other hand, insurance may also be viewed as a method of risk sharing in which a group of insureds collectively bears the risk that a group member will suffer a loss. A particularly intense conflict may arise if losses vary according to the race, sex, or age of the insured, because classification on the basis of these characteristics is likely to be especially suspect.[4]

In short, attitudes toward insurance always seem to be pulling in two directions—one that highlights the risk assessment, or efficiency-promoting features of insurance classification, and the other that stresses insurance's risk-distributional function. This article examines the ways to resolve this tension, by analyzing in detail the competing principles in this field.

Before embarking on this task, it will be useful to place the problem in the context in which risk classification occurs. Part I examines this context. Part II then analyzes the economic forces that affect risk classification. It explains why classification and differential pricing have developed in response to these forces, examines the criteria that may be used to unpack the notions of efficiency and risk-distributional fairness in this setting, and analyzes the problems insurers encounter in setting prices. Next, Part III explores in detail the efficiency concerns that are relevant to classification and pricing systems, and Part IV scrutinizes the differing demands that may be made on classification systems in the name of fairness. Finally, having dealt with the "unfairness of efficiency" in Part IV, the article addresses the coordinate question of the costs of fairness in Part V.

I. The Context

Insurance operates best in the face of a very special sort of uncertainty.[5] The tension between risk assessment and risk distribu-

---

[4] See, e.g., H.R. Rep. No. 100, 98th Cong., 1st Sess. (1983) (proposed prohibition of discrimination in insurance on the basis of race, color, religion, sex, or national origin); S. 372, 98th Cong., 1st Sess. (1983) (same).

[5] When ignorance about the future exists, the transfer of risk resembles gambling more than insuring, because insurers have no way to price coverage. When complete knowledge of

tion is so characteristic of the operation of insurance that risk-sharing schemes from which the tension is missing seem only to resemble what we think of as insurance. For example, if we knew precisely how many losses of a certain sort would occur, but nothing about who would suffer them, then insurance against such losses would be feasible. It would be insurance embodying only the distribution of risk among those insured, however, with no assessment of the extent of each individual's risk: each insured would pay the same premium. Similarly, if we knew who was at risk of suffering a loss if it occurred, but nothing about how many of those at risk would suffer losses, insurance would also be feasible: insureds would be charged retroactively for their proportionate share of whatever losses ultimately occurred. Again, this arrangement would embody risk distribution, but no individualized risk assessment. Although both of these schemes involve risk sharing, each is simpler than the standard insurance arrangement because neither involves any individual risk assessment.

Typically, however, insurers know something about individual risks. Because in such instances it is usually possible both to assess and to distribute risk, the tension between assessment and distribution is inevitable. Risk assessment through classification of insureds into groups posing similar risks necessarily limits the amount of risk distribution achieved by an insurance arrangement, because it uses knowledge about risk expectancies to set different prices for members of different groups. No risk classification system, however, can classify and price individual risks with anything near complete accuracy; the future is too uncertain for that. Nevertheless, when reasonably accurate risk assessment is feasible, insurance classification can promote economically efficient behavior by encouraging insureds to compare the cost of insurance with the cost of investment in loss prevention that would reduce the sum of these two costs.[6] In contrast, when risk assessment is inaccurate but insurance is still available, inefficient behavior is a likely result. This is the "moral hazard" of insurance.[7]

---

the future exists, there is no risk to insure against—there is certainty.

[6] See infra p. 422.

[7] This article uses the term nonpejoratively to refer not only to deliberately unsafe action but also to less, inadvertent, or any other nondeliberate activity that is less safe than it would be in the absence of insurance. For discussions of moral hazard, see Ehrlich & Becker, Market Insurance, Self-Insurance and Self-Protection, 80 J. Pol. Econ. 623 (1972); Holm-

Often a classification scheme can be made more efficient. As we shall see, however, promoting efficiency through risk classification sometimes requires sacrificing other values.[8] The burdens of inaccuracy may be unevenly distributed; risk classes may be based on variables not within the control of insureds; and certain variables may have unacceptable social or moral connotations. For convenience, Part IV describes these concerns as sacrifices in "risk-distributional fairness." Although these concerns are similar in that each sometimes demands inefficient forms of risk classification, risk-distributional fairness itself is not a monolithic notion. An important part of my argument is that we need to be careful to isolate the different kinds of risk-distributional concerns that different approaches to insurance classification raise. Without attention to the differences among these concerns, proposed remedies will be overbroad, undereffective, or both.

A variety of legal tools is available for addressing these issues and regulating the combination of efficiency and risk distribution reflected in the insurance market's classification practices. Legislatures can exercise control through statutes governing insurance classification[9] or through general prohibitions against various forms of discrimination.[10] Legislation in almost every state also delegates considerable authority to insurance commissioners to regulate risk classification and premium rates. Often these commissioners are required to assure that premium rates are not "excessive, inadequate, or unfairly discriminatory."[11] This mandate affords commissioners broad discretion to fashion compromises between the twin goals of efficiency and risk-distributional fairness. Finally, the courts also play a role through judicial enforce-

---

strom, Moral Hazard and Observability, 10 Bell J. Econ. 74 (1979); Shavell, On Moral Hazard and Insurance, 93 Q.J. Econ. 541 (1979).

[8] See infra pp. 428-29.

[9] See R. Keeton, Basic Text on Insurance Law 557-67 (1971); Kimball, The Purpose of Insurance Regulation: A Preliminary Inquiry in the Theory of Insurance Law, 45 Minn. L. Rev. 471 (1961).

[10] For example, in City of Los Angeles Dep't of Water & Power v. Manhart, 435 U.S. 702 (1978), and Arizona Governing Comm. for Tax Deferred Annuity & Deferred Compensation Plans v. Norris, 463 U.S.1073 (1983), the Supreme Court held that anti-discrimination provisions of the Civil Rights Act of 1964 applied to employer-sponsored pension plans.

[11] See, e.g., Cal. Ins. Code § 1852 (West 1972); Ga. Code Ann. § 33-9-4 (1982); N.Y. Ins. Law § 186 (McKinney 1982); Wis. Stat. Ann. § 625.01(2)(a) (West 1984); R. Keeton, supra note 9, at 565.

ment of statutory standards and through oversight of administrative action.[12] In short, the inevitable tensions between risk assessment and risk distribution create the context in which the institutions that make law governing risk classification operate. The next part examines the classification process itself.

II. The Nature of Insurance Classification and Pricing

A. The Emergence of Risk Classification

The starting point for any analysis of insurance classification is an obvious but fundamental fact: insurance is only one of a number of ways of satisfying the demand for protection against risk. With few exceptions, insurance need not be purchased; people can forgo it if insurance is too expensive. Indeed, as the price of coverage rises, the amount purchased and the number of people purchasing will decline. Instead of buying insurance, people will self-insure by accumulating savings to serve as a cushion in the event of loss, self-protect by spending more on loss prevention,[13] or simply use the money not spent on insurance to purchase other goods and services.[14] An insurer must compete against these alternatives, even in the absence of competition from other insurers.

One method of competing for protection dollars is to classify potential purchasers into groups according to their probability of loss and the potential magnitude of losses if they occur. Different risk classes may then be charged different premiums, depending on this *expected* loss.[15] Were it not for the need to compete for protection dollars, an insurer could simply charge each individual a premium based on the average expected loss of all its insureds (plus a margin for profit and expenses), without incurring classification costs.

An insurer can capture protection dollars by classifying because, through classification, it can offer low-risk individuals lower prices. Classification, however, involves two costs. First, the process of

---

[12] See, e.g., Shavers v. Kelley, 402 Mich. 554, 267 N.W.2d 72 (1978); State Comm'r of Ins. v. North Carolina Rate Bureau, 300 N.C. 381, 269 S.E.2d 547 (1980).

[13] I use the term loss prevention to refer not only to safety precautions, but also to changes in activity or production levels that will affect the occurrence of losses.

[14] Cf. Erlich & Becker, supra note 7, at 624 (demand for insurance no different from demand for other goods and services).

[15] For discussion of the concept of expected loss, see Stanford Research Inst. Int'l, The Role of Risk Classifications in Property and Casualty Insurance: A Study of the Risk Assessment Process 40-46 (Final Report 1976).

classification is costly. Insurers must gather data and perform statistical operations on it; marketing may also be more costly when prices are not uniform. Second, classification necessarily raises premiums for poor risks, who purchase less coverage as a result. In the aggregate, classification is thus worthwhile to an insurer only when the gains produced from extra sales and fewer pay-outs outweigh classification costs plus the costs of lost sales.[16] Even in the absence of competition from other insurers, an insurer who engages in at least some classification is likely to capture more protection dollars than it loses.

When there is not only competition for available protection dollars, but competition among insurers for premium dollars, the value of risk classification to insurers becomes even clearer. The more refined (and accurate) an insurer's risk classifications, the more capable it is of "skimming" good risks away from insurers whose classifications are less refined. If other insurers do not respond, either by refining their own classifications or by raising prices and catering mainly to high risks, their "book" of risks will contain a higher mixture of poor risks who are still being charged premiums calculated for average risks. These insurers will attract additional poor risks, and this resulting adverse selection[17] will further disadvantage their competitive positions.

B. How to Evaluate a Classification System

The central concept in a risk classification system is the notion of expected loss—the predicted probability that an insured will suffer a loss multiplied by the predicted severity of the loss. In constructing risk classes, the insurer's goal is to calculate the expected loss of each insured, and to place insureds with similar expected losses into the same class, so that each may be charged the same rate.[18]

---

[16] See L. Freifelder, A Decision Theoretic Approach to Insurance Ratemaking 124 (1976).

[17] Adverse selection occurs when a "group of potential insureds [is] treated alike," regardless of characteristics that might distinguish their expected losses. "A disproportionately high percentage" of adverse risks may then apply for coverage because these insureds will get a "better bargain" than low-risk applicants. R. Keeton, supra note 9, at 8.

[18] The creation of insurance classes produces comparative risk predictions, or risk relativities. Once the classes are fixed, decisionmakers know how the expected losses of one class compare with the expected losses of others. They have not yet, however, set premiums—they have only fixed the proper ratio of the premiums to be charged each class.

Expected loss is a prediction of an insured's actual losses. For two reasons, however, actual losses vary from expected loss. First, calculations of expected loss normally cannot include all relevant variables. A classification based on one or a few variables is likely to be at best a rough estimate of any individual insured's actual loss. If different variables were considered, an individual's expected loss might be entirely different. Second, expected loss is only the predictable component of any individual's *actual* loss. For practical purposes, a large percentage of the actual losses of most individuals and enterprises occur randomly. "Random" losses are either impossible to predict given current knowledge or are too

---

To set actual prices, cost calculations must be made. In practice the process is complex, but in schematic terms it is straightforward. Rates are set by adding together expected losses, expenses, (for example, commissions, administrative expenses, and anticipated litigation costs), a premium to cover the risk of variation from the mean expected loss (if the variation is likely to be volatile), and a margin for profit. See H. Denenberg, R. Eilers, J. Melone, & R. Zelten, Risk and Insurance 509-13 (2d ed. 1974) [hereinafter cited as H. Denenberg]. This calculation is simple enough, although obtaining accurate data to make it is no easy accomplishment.

Decisionmakers may revise the rates charged in either of two ways. See id. at 528-30. The *loss ratio* method compares the actual losses suffered during the policy period to premiums collected and adjusts premiums depending on how this ratio compares to the ratio of expected losses to premiums. If expected losses were projected to be 60% of premiums, but the actual loss ratio turned out to be 80%, then the loss ratio method would suggest a rate increase of 33⅓%, because actual losses had been one-third higher than expected. In taking this approach the loss ratio method makes no change in the composition of risk classes. It merely revises rates from a pre-existing base.

In contrast, the *pure premium* method actually calculates rates anew rather than revising them. The pure premium is the ratio of actual dollar loss to the total number of "exposure units," or units of risk. In workers' compensation, for example, an exposure unit is $100 of salary; so the pure premium for workers' compensation insurance is a rate per $100 of salary paid. The pure premium method yields a new rate to be charged if past loss experience is completely credible, rather than a percentage change in the previous rate.

Because past loss experience is not always statistically reliable, however, the loss ratio and pure premium methods count both actual losses and expected losses in adjusting premiums, thus discounting the predictive value of actual loss experience to some extent. The pure premium method suggests a new premium, however, whereas the loss ratio method suggests only a percentage change in the old premium. The pure premium method therefore makes assigning insureds to new classes more feasible. See Kulp, The Rate-Making Process in Property and Casualty Insurance—Goals, Techniques and Limits, 15 Law & Contemp. Probs. 493, 500-01 (1950).

When there is no reason to think that the variables supporting existing classifications have become less reliable or that use of more powerful variables has become feasible, the two methods of rate revision may be equally satisfactory. Where such changes are possible, however, the pure premium method of revision makes it easier to recognize any increasing heterogeneity in existing classes.

costly to predict. Consequently, even very refined risk classification systems are imperfect predictors. The amount of their inaccuracy depends on the proportion of losses that are random.

The tension between risk assessment and risk distribution in classification is best understood by examining the role that expected loss plays in the five important features of classification schemes. The first three features—the *separation*, *reliability*, and *incentive value* of risk classes—are associated with risk assessment and its concomitant effect on economic efficiency. Two additional features of risk classes—their *homogeneity* and *admissibility*—bear on the risk distributional fairness of the system that uses the classes.[19] These five features reflect the different aspects of economic efficiency and risk-distributional fairness that are relevant in evaluating risk classification systems.

1. Separation

The first key to accurate risk assessment is achieving "separation" between risk classes. Separation measures the degree to which insureds in different risk classes have different expected losses. The greater the separation between classes, the greater the statistical support for charging members of the classes different premiums and the lower the risk of misclassification.

To achieve separation, the difference between the mean expected losses of any two classes should be significant enough to warrant charging them different premiums. Separation will not be pursued, however, beyond the point where the costs of gathering data needed for further refinement exceed the competitive benefit that can be derived from that refinement. A class must be large enough, moreover, for the inferences drawn from its loss experience to be statistically sound, or "credible."

For example, suppose that age and sex correlate with driver accident rates and that both are used in constructing risk classes. If male and female drivers age nineteen to twenty-six are placed in

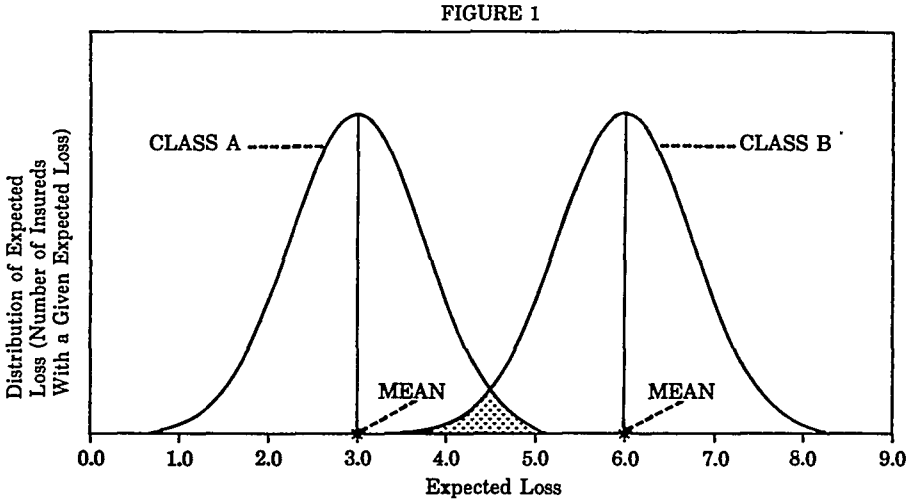
---

[19] See National Ass'n of Ins. Comm'rs, D-3 Advisory Comm., Private Passenger Automobile Insurance Risk Classification 326-27 (1979); S. Rea, Jr. & M. Trebilcock, Rate Determination in the Automobile Insurance Industry in Ontario: The Use of Age, Sex and Marital Status as Rating Variables: A Report Submitted to the Insurance Bureau of Canada 40-41 (1981); Shayer, Driver Classification in Automobile Insurance, *in* Massachusetts Div. of Ins., Automobile Insurance Risk Classification: Equity and Accuracy 3-6 (1978).

separate classes because the males' expected losses are higher, there may not be complete separation.[20] Some twenty-six year-old males may have lower expected losses than some nineteen year-old females. The less overlap of this sort that occurs, the greater the separation between the two classes. In some classification schemes, however, the demands of separation have been sacrificed to assure credibility. Statistical techniques for reducing credibility concerns are available, but they pose accuracy problems of their own.[21]

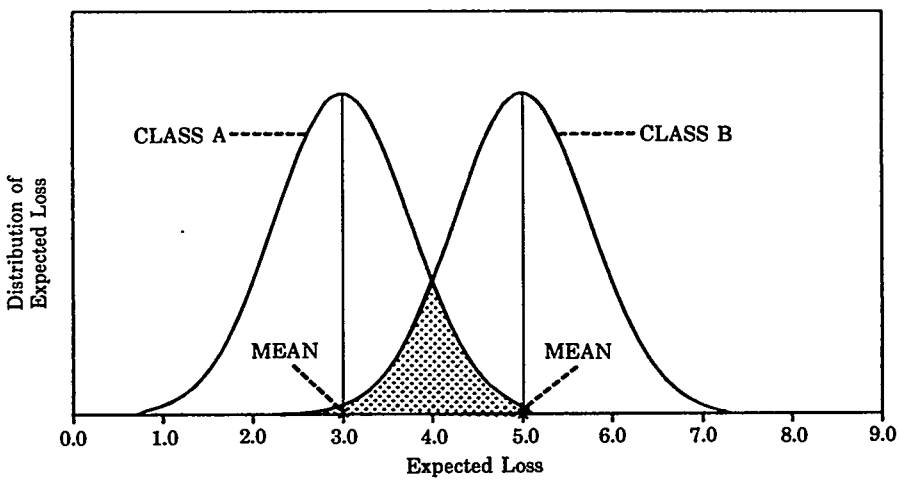
Figures 1 and 2 illustrate two different degrees of separation. The vertical axis measures the distribution of expected loss (the number of insureds with a given expected loss) and the horizontal axis measures the amount of expected loss. The difference between the mean expected losses of classes *A* and *B* is greater in Figure 1 than in Figure 2. Thus, the separation between the classes in Figure 1 is greater than that in Figure 2. The shaded area of overlap between the two classes indicates the portion of all insureds who have similar expected losses but who are charged different premiums. The smaller overlap in Figure 1 reflects the greater separation between the two classes.

FIGURE 1



[20] See Stanford Research Inst. Int'l, supra note 15, at 50-51.
[21] See id. at 78.

FIGURE 2



2. Reliability

A second aspect of accurate risk assessment (and therefore of efficient classification) is the reliability of the characteristics on which risk classes are based. Not only is separation between classes desirable in theory, but classification should be susceptible to as little administrative error and fraud as possible. The main reason risk classes may be unreliable is that insurers sometimes cannot easily verify the data that applicants furnish. Insurers may require medical examinations before accepting life insurance applications, use driver's licenses to establish age, and rely on data banks to verify accident experience. When verification would be difficult, however, the insurer must choose between disregarding a potentially predictive variable and undermining the reliability of its classifications. For example, mileage driven is a comparatively good predictor of driver accident rates, but is often considered too unreliable a variable to warrant its use.[22] Smoking apparently accounts for a sizable minority of residential fires,[23] but a non-

[22] See Shayer, supra note 19, at 17.

[23] See O'Hare, Information Strategies as Regulatory Surrogates, *in* Social Regulation: Strategies for Reform 221, 232 (E. Bardach & R. Kagan, eds. 1982).

smoker's discount for purchasers of homeowner's insurance is rare. In each of these instances the need for reliability prevents the insurer from using a potentially powerful classification variable.[24]

3. *Incentive Value*

Risk classifications should reflect differences in expected losses between classes of insureds; ideally, they should also create loss prevention incentives for insureds. To achieve this goal, classes should (insofar as possible) be based on variables within each insured's control. In this context "control" encompasses not only the ability to conduct activities more safely, but also the capacity to vary levels of activity or production to reduce or prevent losses. Insurers should also be able to reclassify insureds periodically to account for the changes in expected loss resulting from insureds' prevention efforts.

There are two major ways to use classifications to influence loss prevention behavior:[25] feature rating and experience rating.[26] Fea-

---

[24] The nature of the coverage can also affect the reliability of the classifications it employs. For instance, there are two major kinds of liability insurance policies. *Occurrence* policies provide coverage against liability arising out of acts of the insured occurring during the policy period, no matter when a claim is eventually lodged against the insured. Pricing occurrence coverage requires prediction of claims that will be made in the future. Yet the past is not necessarily a very reliable gauge of the future, especially where demographics and claims consciousness are changing. Thus, in fields where claims may not be brought for many years—medical malpractice or toxic torts, for example—pricing occurrence policies is highly speculative because expected losses are hard to predict.

Claims-made policies, on the other hand, cover the insured against all claims that are made during the policy period, regardless of when the activity giving rise to the claim occurred. Claims-made pricing requires much less prediction of the future because only the claims that will be filed during the forthcoming policy period need be predicted.

The difference between the two kinds of policies is that the insurer bears the risk of an uncertain claims future under an occurrence policy; a claims-made approach shifts much of that risk back to the insured. In effect, the claims-made approach reflects insurers' interest in retaining the ability to revise rates upwards as claims experience evolves. That ability is important to the insurer where its confidence in predictions of expected loss is unavoidably low. The result, however, is that claims-made insureds have less real insurance, because they do not have coverage against any claims to be filed after the close of the current policy period. Because they must wait until future policy periods to purchase such coverage, claims-made insureds bear much of the risk of an uncertain claims future—a risk against which occurrence coverage would protect them.

[25] See H. Denenberg, supra note 18, at 516-22.

[26] This form of classification is sometimes also called loss rating. For a discussion of feature rating and experience rating, see N. Doherty, Insurance Pricing and Loss Prevention 48-53 (1976).

ture rating relies on observable features of the insured or the insured's operations to calculate expected losses and construct risk classes. The features used as variables may be levels of safety or levels of activity. Basing fire insurance rates on building characteristics (such as size, materials used in construction, and presence or absence of sprinkler systems) and basing product liability insurance rates on sales volume are both examples of feature rating. Data collected over the years, or insurers' intuitive hunches,[27] suggest which features are correlated with loss rates and should therefore be used as classification variables.

Experience rating, on the other hand, uses the loss experience of the insured during one period to help set the premiums charged in the following period. Experience rating is usually used in combination with feature rating. For example, drivers may be classified initially according to their sex, age, and marital status; their rates then may be revised upward or downward in subsequent policy periods, depending on their claims experience.[28]

Feature and experience rating affect loss prevention incentives in different ways. Over a given policy period, feature rating creates no loss prevention incentives, but in the long run this form of rating can have incentive effects. If the insured has not made a large, fixed investment in the features in question, he may he able to modify them during later policy periods. Conversely, feature rating will have little incentive effect even in the long run when the risk insured against depends mainly on features that are fixed over the long term.[29]

The long-term effects of feature rating therefore depend on the kind of control the insured can exercise over the features used as classification variables. Some types of feature rating may affect activity levels, but not the safety of given activities. For example, a family that decides not to purchase a second car may effectively

---

[27] I employ the term "hunch" deliherately. Sometimes fairly refined categories are used without much data to back them up. Little statistical support seems to exist, for example, for many of the credits and charges made for the supposedly good and bad features of a fire risk. See H. Denenberg, supra note 18, at 533.

[28] A third form of rating, "retrospective" rating, functions much like experience rating. It calculates premiums only after the expiration of the policy period. Rates are based on the losses suffered during the period, subject to a specified minimum and maximum.

[29] See Shavell, supra note 7, at 551 (ex ante rating may be optimal where expenditures on risk-posing assets are fixed over a period; ex post rating may be preferable where risk-reducing activity can be varied).

limit the amount of its teen-age driving, but not the safety of the teen-age driving still allowed. Similarly, over the long term, fire insurance rating based on building materials may affect the number and types of houses that are built, but it cannot directly affect residential fire safety activities.

Unfortunately, some feature rating schemes have little incentive value even when they are based on controllable variables. A product manufacturer has control over its sales volume. Feature rating based on volume, however, is likely to have little effect on safety because the insured has no incentive to reduce sales to lower premiums. Still other forms of feature rating can affect levels of safety if they are based on modifiable features of an activity. Examples of such features include the use of a comparatively safe kind of equipment or vehicle or the installation of sprinkler or alarm systems.

Interestingly, both safety levels and activity levels can be indirectly affected even when noncontrollable features are used as classification variables. Feature rating based on noncontrollable characteristics, such as age or sex, can affect activity levels by encouraging insureds to reduce the level of their involvement in the insured activity. Moreover, so long as other features of the insured's operations affecting risk are controllable, safety levels can change if insureds are encouraged to reduce their reliance on insurance and to shift instead to safety measures for protection against the risk in question. Ideally, it may be preferable to move to experience rating under these circumstances, although such a move may not always be feasible.

Experience rating has different advantages and shortcomings. The insured who is subject to experience rating has a decided incentive to reduce his claims or liability experience because his future premiums will be affected by that experience. Unlike feature rating based on controllable variables, however, experience rating does not indicate which behavior the insured should alter to modify his expected loss. Instead, the insured is given an incentive to discover a method for himself. Further, when past losses are poor predictors of future losses, experience rating may create incentives to alter behavior that is not closely related to losses.

In terms of incentive value, therefore, feature rating is probably preferable when the effects of particular variables on expected loss are relatively predictable and when changes in other conduct prob-

ably will not significantly alter loss experience. Experience rating, on the other hand, is probably preferable when the effect of different variables on loss varies from insured to insured, or when insureds are in a better position than the insurer to determine what features of their activities can be most effectively varied to affect loss experience. Feature rating provides insureds with the insurer's conclusions about the best way to reduce losses; experience rating reflects the conclusion that the insured, not the insurer can best make this determination.[30]

The incentive value of both feature and experience rating also depends, however, on the information that rates and rate changes communicate to insureds. Even otherwise appropriate feature rating categories and finely-tuned adjustments in premiums through experience rating can have little effect on safety incentives if insureds do not know of the potential insurance savings available from changes in their operations or reductions in their loss experience. In this sense the incentive value of rate structures depends not only on pricing in accord with expected loss, but also on the simplicity and explainability of rating variables and price adjustments.[31]

Much products liability insurance, for example, is provided by comprehensive general liability policies that afford the insured coverage against a variety of risks. Because in the past the price of the products liability component often was not separately indicated, it was difficult for insureds to ascertain what portion of any

---

[30] In a sense, then, the differences in the incentive effects of feature and experience rating in insurance resemble certain differences between negligence and strict liability in the law of torts. Like negligence liability, feature rating imposes a cost on the insured for characteristics that are correlated with the occurrence of losses. The insured then decides whether changing those characteristics that can be changed will be worth the reductions in feature rates these changes would produce.

Experience rating, on the other hand, imposes a cost on the insured that is dependent only on the occurrence of a loss. In this sense it resembles strict liability. If the experience-rated insured is in a better position than the insurer to determine what features of behavior might be altered to reduce the sum of the costs of insurance and loss prevention, then this form of rating will be economically superior to feature rating. For a theory of the difference between negligence and strict liability, See Calabresi & Hirschoff, Toward a Test for Strict Liability in Torts, 81 Yale L.J. 1055 (1972). Another parallel is that experience rating, like strict liability, may have more impact on activity levels than fault-based liability or feature-based risk classification. See Shavell, Strict Liability Versus Negligence, 9 J. Legal Stud. 1 (1980).

[31] See H. Denenberg, supra note 18, at 516; N. Doherty, supra note 26, at 54-55.

premium adjustment could be ascribed to changes in expected products liability losses.[32] In short, not only are costs involved in generating the information on which rate classes are based and in communicating it to insureds; there is also a cost to the insured of consuming that information.[33] The more complicated or confusing the information, the greater the cost of consumption, and the less incentive value any form of rating is likely to have.

4. Homogeneity

The fourth aspect of any risk classification system is its homogeneity. Social concerns operate here instead of economic incentives. Because all members of a class are charged the same rate, it is desirable that they have similar expected losses. The more homogeneous the class membership in this respect, the stronger the argument for charging each member the same rate. Like insureds are then treated alike. On the other hand, the more varied the expected losses of the insureds in a class, the weaker the argument for the classification.

Figure 3 illustrates the concept of homogeneity. Class *A* is more homogeneous than class *B* because the expected losses of its members are more similar than those in class *B*. In other words, the average deviation from the mean expected loss in class *B* is greater than in class *A*.

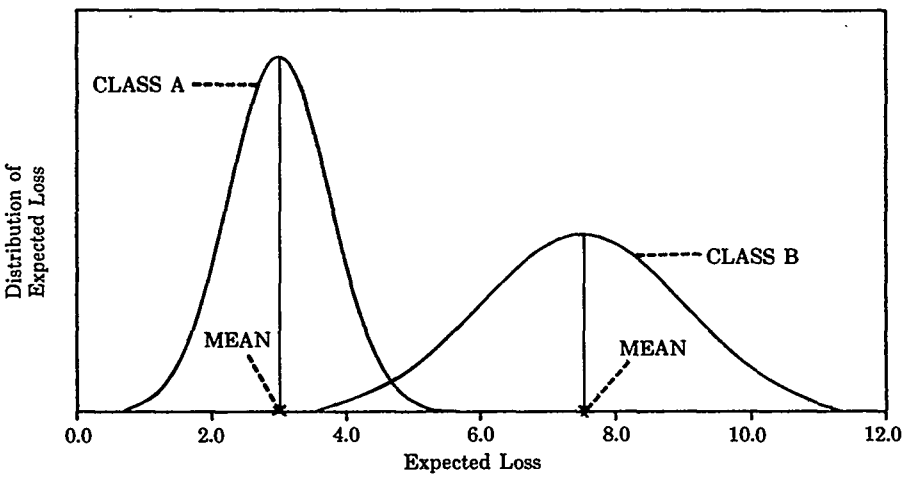
Like separation, homogeneity is a relative notion. No risk class is completely homogeneous, because a class must be large enough to support statistically sound predictions about its claim experience. Moreover, even in an apparently homogeneous class the insureds might have very different expected losses if a different variable were used for classification. For example, suppose that a twenty-year-old male driver is twice as likely as a forty-year-old male driver to be involved in an accident. Separate rate classes based on age might then be worthwhile. Nonetheless, data on other predictive characteristics, such as miles driven, could be unavailable, unreliable, or prohibitively expensive to obtain. The variable used to

---

[32] See 6 U.S. Dept. of Commerce, Interagency Task Force on Product Liability 197-99 (Final Report 1977).

[33] For a short discussion of the costs of consuming information, see O'Hare, supra note 23, at 228-29.

FIGURE 3



establish the classes would therefore explain only a portion of the losses of those classified. The remainder of their expected loss would have to be considered uncertain.[34]

Nevertheless, there may be several valid reasons for a classification system based on variables that explain only a portion of the losses of members of the established risk class. First, such classification may approach the practical limits of predictive capacity. Second, when several factors are combined into multi-variable classes, their predictive value is higher and classes are more homogeneous. Third, objections to seemingly heterogeneous classifications may be based as much on the nature of the variables used or ignored in the classification as on the heterogeneity they create. Life insurance and annuity rates, for example, are based largely on the life expectancies of individuals at the insured's age. This heavy reliance on age as a classifying factor in life insurance is rarely sub-

[34] "Uncertainty" in expected losses refers to the difference between an (assumed) individual expected loss for each insured and the expected loss that the classification system actually assigns to the insured. Because one cannot measure much of an individual's expected loss, "it is possible for a policyholder in a high-risk class to have an expected loss (if it could be known exactly) lower than the average of lower risk classes, and vice versa." National Ass'n of Ins. Comm'rs, supra note 19, at 142; see also id. at 156 (questioning the utility of assuming that there is an "individual" expected loss that cannot be measured).

jected to objections as strong as those leveled against heavy reliance on age in classifying automobile insurance risks.[35] The difference in attitudes may be related more to the belief that actual loss experience or some specific characteristic within the insured's control would be a preferable variable for use in automobile insurance than to any perceived difference in the predictive power of age in the two settings. In short, the demands made in the name of homogeneity may depend on whether the available alternatives are superior in other respects, such as incentive value or admissibility.

Finally, homogeneity should be understood not only as a relative notion, but also as a comparative notion. Risk classes can be evaluated for homogeneity not only in relation to an ideal, but also in comparison with one another. An otherwise acceptable degree of heterogeneity in one class could become objectionable because of the greater homogeneity of other classes. This difference in homogeneity would place a greater risk-sharing burden on members of the more heterogeneous class than on those in other classes. As we shall see when we address risk-distributional issues more explicitly, this possibility may lead to classification schemes designed to avoid uneven distribution of such burdens.[36]

5. *Admissibility*

Some variables with predictive power may be socially, legally, or morally inadmissible for use in constructing risk classes. For example, race, sex, or age may be good predictors of certain kinds of loss experience. Race, however, is almost always an inadmissible consideration, and both sex and age are sometimes objectionable, depending on the kind of insurance involved.[37] In addition, characteristics that are admissible on their face but that have a disparate

---

[35] See, e.g., Stone, Findings and Decision on 1978 Automobile Insurance Rates, *in* Massachusetts Div. of Ins., supra note 19, at 180-81 (arguing that age as a variable in the automobile insurance rating system serves only as a proxy for variables more causally related to accidents, such as driving experience).

[36] See infra text accompanying notes 64-76.

[37] Certain kinds of age discrimination, for example, may be acceptable because the elderly benefit from the classification. For example, if elderly drivers and their passengers have below-average losses when they are involved in automobile accidents because their incomes are below average, then charging them less for first-party no-fault insurance may be quite acceptable. In this setting, age is an admissible variable. Charging the elderly higher rates for liability insurance, however, (assuming that their loss experience would warrant higher charges) might well be unacceptable.

impact on the members of one race or sex also may be suspect. Most automobile insurance classifications use not only age and sex as variables, but also the insured's territory of residence. Because the inner cities are high-risk areas, young, black males tend to be charged very high rates. Even when the impact of these variables is not so pronounced, the use of a territorial variable tends to have income-regressive effects because a higher proportion of the poor live in urban areas, where accident rates are high, than in the suburbs, where accident rates are lower.

The point to underscore here is that admissibility concerns can veto, or at least call into question, variables that otherwise might satisfy one or more of the four classification criteria discussed above. In this sense admissibility is not simply a separate way of assessing a classification, but a conclusion about the validity of a particular form of classification in light of standards external to the classification system.

No system of classification can simultaneously maximize satisfaction of each of the five classification criteria. Variables that on paper produce homogeneity and separation may not be reliable in practice. Yet those that are reliable may produce less homogeneous or less separated classes. Variables not within the control of insureds may produce the most homogeneous classes, but may sacrifice incentive value. Finally, inadmissible characteristics may occasionally produce the greatest separation or homogeneity. As a result, all systems leave themselves open to the criticism that any particular goal is not being fully achieved. In a system of mixed goals, however, this impossibility of perfection is not a prescription for unresolvable conflict or unprincipled compromise. Instead, it is an invitation to weigh the strength of each goal and to fashion a system that most closely reflects this weighting. We now turn to examine directly the risk assessment and risk distribution goals of insurance classification.

III. Risk Assessment and Economic Efficiency

This part discusses the cluster of economic values that the legal literature tends to see as "efficiency" concerns. There are imprecisions inherent in speaking this way, because efficiency is not a

monolithic notion, especially in the technical literature.[38] Nevertheless, it is useful to rely on the general notion of efficiency in the production of economic value and to distinguish that notion from concern with the fair distribution of value—the subject of Part IV. To understand the nature of efficient risk classification, we need to use the criteria previously developed in order to isolate both the positive effects of classification and the limits on their complete achievement.

A. The Nature of Efficient Classification

Other things equal, insurers strive to charge insureds in accord with expected costs,[39] which equal their expected losses plus a portion of the other costs of providing coverage.[40] To the extent that risk classes and prices conform to this standard, a number of results follow. The first effect is that individual insureds pay premiums based on expected losses and thereby share the risk of random losses.[41] The members of each class are charged in accord with their expected costs, so that total premiums cover the aggregate losses of the class. No subsidies run from one risk class to another. The only subsidy under this ideal flows from the lucky members of the class to the unlucky.

Even efficiently classified insurance coverage, therefore, has elements of both risk assessment and risk distribution, but the scope of each is distinct. Individual insureds are assessed the risk of suffering expected losses and are charged on that basis. The risk of suffering random losses is distributed among all insureds.

The second effect of an efficient classification system is that it does not discourage insureds from allocating an optimal amount of

---

[38] For a discussion of the different senses in which the concept of "efficiency" can be employed, see Coleman, Efficiency, Utility and Wealth Maximization, 8 Hofstra L. Rev. 509, 512-20 (1980).

[39] Technically speaking, an efficient system would charge in accordance with marginal expected costs—the additional cost added by each insured purchasing coverage. Because most costs in this context, however, are marginal costs—for example, the amount of the loss an insured suffers and the costs of defending him against suit—the simplification seems warranted. See Williams, Price Discrimination in Property and Liability Insurance, *in* Insurance, Government and Social Policy 210-12 (S. Kimball & H. Dennenherg eds. 1969).

[40] In the discussion that follows, I use the terms "expected loss" and "expected cost" interchangeably, except where it is important to note the difference.

[41] See L. Freifelder, supra note 16, at 70; National Ass'n of Ins. Comm'rs, supra note 19, at 178.

resources to loss prevention. Because insurance is priced in accord with expected cost, insureds have the incentive to compare the cost of protecting against risk through insurance with the cost of reducing risk through loss prevention. Efficient classification discourages insureds from purchasing insurance when they can more cheaply protect against risk by investing in loss prevention.[42] In contrast, inefficient classification may produce suboptimal loss prevention incentives. When coverage is priced below expected cost, for example, insureds may not take safety precautions that would otherwise be worthwhile. In this situation, they can obtain equivalent protection against risk by purchasing insurance at a lesser cost than the precautions.

Finally—and this is a vital point—an efficient classification system does not strive to make its prices equal expected cost when improvement in accuracy is not worth the cost of achieving it. Information about expected cost is accumulated and risk classes are thereby refined only so long as the competitive benefits of refinement are worth their cost. Consequently, efficient classification does not recognize all individual differences. When an insurer can no longer attract or make enough profit from additional low-risk insureds to justify discovering and classifying them, an equilibrium is reached and no further refinement occurs. Some groups may then seem to "subsidize" or be "subsidized" by others.

For example, suppose that people raised on farms are especially poor drivers or that obstetricians born in Ohio are unusually immune to malpractice suits. Because classification systems are unlikely to have the information necessary to make these variables the basis of risk classes, neither farm-born drivers nor Ohio-born obstetricians will be charged exactly in accord with their true expected costs. Thus, the former may seem to be subsidized by other drivers, and the latter may seem to subsidize other insured physicians. It is a bit misleading, however, to say that a situation in-

[42] The description here actually is a bit over simplified, although it is accurate enough for our purposes. Insurance can provide coverage against a specified amount of loss, or it can include deductibles or coinsurance tbat protect against only a portion of any loss. Similarly, some loss prevention activities reduce risk across-the-board, whereas others may reduce the risk of large loss but have little effect on the risk of small losses, or vice versa. The combination of insurance and loss prevention adopted by any insured will depend not only on the comparative costs of each kind of investment in protection against risk, but also on the nature of that protection and the nature of the individual's risk aversion that demands it.

volves a subsidy even when it would be inefficient to make the investment necessary to discover and eliminate the "subsidy."[43]

This example makes it plain that there is nothing special or preordained about the classifications that turn out to be efficient. Had insurers begun decades ago to maintain data about farm-born drivers or the birth places of obstetricians, it might now be efficient to use these variables for classification. In many cases, however, a new approach cannot be taken without sacrificing real economies. Even if restructuring a classification system would otherwise be efficient, probably no individual insurer would have an incentive to restructure. Competitors would take advantage of the classifications introduced by the innovating insurer and compete on an equal basis for the newly-discovered low-risk insureds, but without having made the investment required of the innovator.[44] Some form of collective action would therefore be required for the innovation to occur. We will return to this problem and its possible solution in Part V.

To sum up the implications of the discussion thus far, insurance relies on group rather than individual estimates of expected loss. With few exceptions (such as large enterprises with detailed loss histories and frequent current losses), estimating expected loss "individually" is impossible. Most individual loss experience is not statistically credible enough to warrant individual rating, though a few insurers occasionally gamble on unique risks—the well-being of John McEnroe's left arm, for example. Group probabilities provide the credibility necessary to the predictions that are at the heart of the insurance system. Until an individual insured is treated as a member of a group, it is impossible to know his expected loss, because for practical purposes that concept is a statis-

---

[43] I have entirely omitted discussion of one other efficiency-promoting feature of insurance, because its operation does not depend on classification: the stability and security provided when insurance protection reduces the uncertainty produced by the threat of future loss. A potential insured's uncertainty about future losses tends to inhibit the productive use of assets, because some assets must he reserved or left liquid to cover potential losses. The purchase of insurance stabilizes the reserving process by fixing a sum—the insurance premium—that is automatically "reserved" to cover losses. Remaining assets (up to a point where policy limits are exhausted) can therefore be used in more productive ways.

[44] It is possible that even if an innovating insurer cannot appropriate all the benefits of its new classification, the costs of implementing it may be so low that they are offset even by the share of the benefits that does accrue to the innovator. Given the costs of compiling data, however, such instances would not be common.

tical one based on group probabilities. Without relying on such probabilities, it would be impossible to set a price for insurance coverage at all. In this sense there is risk sharing even within the risk assessment component of insurance classification. For practical purposes, no individual can have a "true" expected loss. Rather, insureds share the risk that characteristics of those in their risk class, not considered in the classification, render the class's expected loss higher than it would be were those characteristics considered in setting premium rates.

B. The Impact of Separation, Reliability, and Incentive Value on Efficiency

Three of the criteria discussed earlier—separation, reliability, and incentive value—are directly related to the efficiency of a classification system.[45] The greater the potential separation between classes, the more likely it is that the competitive benefit derived from additional refinement will be worth the cost. The more reliable the variables on which classes are based, the more worthwhile the level of refinement, because it will represent classification actually made on the basis of expected loss rather than classification that can be circumvented or confused in practice.

Perhaps the most striking of all the effects of an efficiently structured classification system is its incentive value. Efficient risk classification performs an optimizing function by inducing insureds to compare the cost of coverage with the cost of safety precautions and other forms of loss prevention. If insurance is underpriced, the insured has an incentive to purchase too much coverage and to invest too little in loss prevention. If insurance is priced in accord with expected cost, however, it can help promote optimal investment in insurance and loss prevention.

This incentive for optimal investment exists even when the variables on which risk classes are based are not within the insured's control. The insured still has the incentive to minimize his or her overall cost of protection against risk through safety expenditures (or reductions in activity levels) so long as these expenditures produce greater protection than a similar investment in insurance. When a noncontrollable variable is used to help set the price of

---

[45] See S. Rea, Jr. & M. Trebilcock, supra note 19, at 41.

insurance, reducing loss experience will not reduce insurance cost directly, but reducing *losses* may reduce the amount of coverage needed. For this reason the incentive to optimize loss experience operates even when noncontrollable variables are used. Thus, although the use of noncontrollable variables may sometimes be considered unfair, it is not necessarily inefficient.

Obtaining a related optimizing effect, however, does depend on whether the classification variables used are within an insured's control. Risk classification can perform an informing function by supplying the insured with information about the features of his behavior or operations associated with expected loss. This information indicates the types of safety precautions that the insured should take to reduce expected loss. Knowing better how to invest in protection, the insured will need to purchase less insurance to obtain optimal risk protection. The more accurate the classification, the more useful the information may be. Yet the classification can perform this function only when the variables on which it is based are within the insured's control.

If classification sends messages about features outside the insured's control—age or sex, for example—the messages cannot be converted into action. When such messages are about features within the insured's control, they may assist the insured in allocating his protection dollars more efficiently and thereby reduce his need for insurance. If additional investment in loss prevention serves to reduce the insured's expected loss, they may also help reduce the cost of the future coverage that the insured will continue to purchase.

Even experience rating, which points to no particular feature of the insured's activities that might be altered to reduce loss experience, plays a role in the informing function of insurance. Although experience rating does not indicate specific precautions that would reduce coverage costs (or mitigate cost increases), it indicates the costs that could have been saved this year by avoiding the losses that occurred last year. This figure is a rough estimate of the potential benefit of the expenditures that would prevent similar losses in the ensuing years.

Unfortunately, each of these possible incentive effects is only the reflection of an ideal. At various points in real classification systems, significant slippage occurs. First, the cost of obtaining information about expected loss is high enough in many contexts to

make accurate classification extremely difficult. In such cases the proportion of losses that are effectively random, and therefore broadly distributed, is high. The incentive value of the resulting rates falls far short of the efficiency ideal.[46] As Rothschild and Stiglitz have shown in theoretical terms, the problem of distinguishing high- from low-risk insureds may even make it impossible to achieve the competitive equilibrium characteristic of an efficient insurance market.[47] Continual reshuffling of rates and classifications may be the consequence.

Second, a classification system may fail to achieve all its potential incentive effects because losses may be so infrequent that experience rating can only modestly affect rates. In such endeavors the occurrence of accidents has some predictive value, but the absence of accidents proves little. For example, the average driver has an accident only once every ten to twelve years.[48] Here, as elsewhere, statistical techniques are necessary to determine how much the occurrence of one accident increases the probability that the insured will have another.[49] A low-risk driver may, by chance, have more accidents during a three-year period than a driver with a higher expected loss.[50] Because the occurrence of a loss is only a rough indication of the new expected loss that should be used in resetting a premium, the incentive value of that premium may be crude.[51]

Third, slippage occurs because expense costs (known in the trade as expense "loading") may be calculated in simple proportion to

---

[46] See Stanford Research Inst. Int'l, supra note 15, at 48-49 (risk assessment processes in effect in 1976 explained only 30% of the variance of expected loss distribution among drivers).

[47] See Rothschild & Stiglitz, Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information, 90 Q.J. Econ. 629 (1976).

[48] National Ass'n of Ins. Comm'rs, supra note 19, at 44. The phrase "ten to twelve" reflects different definitions and measurements of "accidents." Id. at 49.

[49] The prevalent technique is use of the "Poisson" model, which helps predict accident likelihood variability. See Stanford Research Inst. Int'l, supra note 15, at 175-79 (Supp. 1976).

[50] See National Ass'n of Ins. Comm'rs, supra note 19, at 44.

[51] The occasional insured loss then will function in much the same manner as the cause in fact requirement in tort law. It will simply be a useful way of calculating some of the costs that the insured should be required to internalize in order to create safety incentives on his part. See Calabresi, Concerning Cause and the Law of Torts: An Essay for Harry Kalven, Jr., 43 U. Chi. L. Rev. 69, 85 (1975).

expected loss.[52] Yet if expected expenses are not proportional to expected loss, the incentive value of the resulting rates will be diluted. For example, a portion of the expenses allocable to commissions and administrative costs is incurred regardless of whether the insured makes a policy claim. Insureds paying high premiums because of high expected losses pay a disproportionate share of such expenses if the latter are assessed in proportion to premiums. Low-risk insureds are correspondingly underassessed. Because the portion of premiums devoted to expenses often is sizeable, the resulting distortions may have a significant impact.

Finally, legislative or regulatory intervention may impede accurate classification. Such intervention may be designed to protect against the use of inadmissible variables or may serve other purposes. Though such intervention may be justified on other grounds, it can reduce the system's efficiency if it hinders accurate classification. Moreover, even in the face of intervention, insurers who are prevented from classifying in the manner they desire still can use underwriting decisions to help them continue pricing in accordance with expected cost. For example, insurers may refuse to sell coverage to applicants whom they are prohibited from charging in accord with expected cost, in order to preserve the integrity of risk classes.[53] As we shall see in Part V,[54] although legal strategies can neutralize such moves, they have costs of their own.

C. Homogeneity and Admissibility

Homogeneity and admissibility are conspicuously absent from the list of criteria relevant to efficient risk classification. Homogeneity is absent because once an optimal amount of separation between classes has been achieved (when the benefit of further separation would not exceed its cost), homogeneity is economically irrelevant.[55] An example will help to illustrate this point.

Suppose that optimal separation already had been achieved by dividing a group of insureds into two risk classes containing a twenty percent overlap (twenty percent of the insureds in each class would have the same expected loss as their counterparts in

---

[52] See National Ass'n of Ins. Comm'rs, supra note 19, at 144-45.

[53] See id. at 16.

[54] See infra text accompanying note 84.

[55] See S. Rea, Jr. & M. Trebilcock, supra note 19, at 66.

the other class). By definition, the cost of further classification to reduce the overlap would be greater than the profits that could be captured or the cost savings that could be returned to insureds by charging more nearly in accord with the expected losses of that twenty percent. The increase in the homogeneity of the classes that would result from further refinement would therefore not be worth its cost.

Thus, separation measures the differences in expected losses between members of different classes. It is the key to achieving accurate classification and efficient pricing. Homogeneity is irrelevant to this goal because it is merely a residual measure of the differences between the expected losses of members within a given risk class. Greater homogeneity is often a by-product of greater separation, because separating and charging different risks differently is likely to result in more homogeneous classes. Once there is optimal separation, however, achieving greater homogeneity will yield no additional economic benefit. This does not mean that homogeneity is an inappropriate concern. That would be true only if efficiency were the exclusive concern of those regulating risk classification. The social desirability of treating like insureds alike by having roughly comparable degrees of homogeneity among classes may make it worthwhile to sacrifice some degree of efficiency.

Admissibility considerations may also conflict with efficiency. The use of any variable that would improve the reliability of or separation between classes would enhance the efficiency of the system, because it would assist in charging insureds in accord with their expected losses. If that variable is inadmissible, the efficiency of the system is undermined. To see the significance of homogeneity and admissibility concerns we must therefore turn to the set of considerations that may warrant a sacrifice of efficiency in classification.

IV. Risk-Distributional Fairness

In addition to asking how well a classification system assesses risk to produce efficient results, we may also raise a series of different questions regarding how insurance classification and pricing systems distribute risk. Because these questions are not always clearly formulated, they need both explication and evaluation. Moreover, because the questions are not all of the same order, it is important to consider the different solutions that are available for

the different kinds of unfairness that these questions highlight. Certain solutions may appropriately be fashioned within the insurance system, but for other risk-distributional problems, noninsurance approaches are more suitable. This part begins the analysis of such solutions by examining the distributional objections to the drive for efficiency in classification. The analysis continues in Part V, which explores the corresponding problem of the costs of satisfying these distributional objections.

Criticisms of risk classification schemes seem to fall into three general categories. The first cluster of criticisms is composed of *accuracy-equity* concerns: demands that classification and pricing closely reflect expected cost, so that low-risk insureds are not forced to subsidize high-risk insureds and so that the burdens of inaccuracy are equitably distributed. Accuracy concerns obviously have much in common with the efficiency notions discussed in the preceding section. Efficiency, however, is a characteristic of an entire sytem; risk-distributional fairness is a notion that also pertains to the treatment of individuals within that system and touches on more than economic considerations alone. The second category of criticisms is composed of *control-causality* concerns. These involve the contention that risk classes should be based on variables that are within the control of or at least caused by the insured. The third set of criticisms is directed at the use of *suspect variables.* These criticisms are occasioned by the use of variables that are "suspect," even apart from accuracy-equity or control-causality issues. A fourth consideration, different from these critisisms, is one in favor of *redistributional policies.* This consideration does not necessarily raise fairness questions in the same way as the first three; rather, it involves separate objectives of public policy.

A. Accuracy-Equity

1. Inaccuracy in General

Criticisms of a classification as being inaccurate or unrefined may raise two quite different issues. The first is whether, as a matter of fairness, people should pay for insurance coverage only in proportion to the cost of what they can be expected to consume. If so, then inaccurate classification improperly overcharges some and undercharges others. This notion appears to have libertarian overtones in its expression of a distaste for the "subsidies" that inhere

in inaccurate classification. Except in the few instances where insurance coverage is legally mandated, however, people are at liberty to decline to purchase insurance if it is priced too inaccurately to suit them. The argument against inaccuracy in general seems therefore also to include the idea that there should be a right to purchase accurately priced coverage, even when purchasing is not required.

The weakness in this argument is that because no risk class is completely homogeneous, the slightly lower risks within a class always seem to subsidize the slightly higher risks.[56] About half of any class, after all, has expected losses below the class average. Further, there is inevitably some misclassification, and some insureds who are classified properly according to the terms of the system would be classified differently under a more refined one.[57] A final and more fundamental weakness is the following: to argue that the risk-sharing inherent in a classification scheme automatically constitutes "subsidization" is to misunderstand the nature of classification.

In a system of market-supplied insurance, insurers have an incentive to classify accurately even in the absence of legal intervention. As we saw earlier,[58] they will stop pursuing greater accuracy only when its marginal cost exceeds its marginal benefit. In a competitive market, not only insurers but also insureds benefit when classification is refined up to but not beyond this point, because the costs of classification are reflected in the price of coverage. If a decrease in the cost of coverage that would result from greater accuracy would be exceeded by an increase reflecting the extra cost of obtaining that accuracy, then it would not be in anyone's economic interest to have generally more accurate classification.

Critics of inaccuracy must then be making one of two arguments. First, when the amount of accuracy in the system is already economically optimal, they must be arguing that everyone should be charged more for coverage simply to assure greater accuracy in general. No one would favor such a system, however, because no one would benefit from it.[59] Everyone would pay more for cover-

---

[56] National Ass'n of Ins. Comm'rs, supra note 19, at 17.

[57] See H. Denenberg, supra note 18, at 518.

[58] See supra p. 410.

[59] Cf. National Ass'n of Ins. Comm'rs, supra note 19, at 17 (proposal for flat insurance

age, unless for some reason the cost of obtaining that extra accuracy were borne disproportionately by different risk classes.

A second argument against inaccuracy can be made when the classification system is inefficient—when it is not achieving optimal accuracy. The argument is that because it is wrong to allocate resources inefficiently, inaccurate classification is unfair to everyone concerned. To the extent that the "waste" entailed in the failure to allocate resources efficiently is morally as well as economically objectionable,[60] this argument has some weight. But it is far from being a claim about individual fairness in risk bearing. Instead, the argument is a conflation of efficiency with morality.

In either case, it is worth underscoring that "inaccurate" classification of any given individual should not itself be objectionable. Any classification system can achieve only a certain level of accuracy. Even if that level has not yet been reached, inefficiency, not unfairness, is properly the focus of criticism.

2. *Differential Inaccuracy*

An entirely different objection to inaccuracy is based on a claim about the unfairness of *uneven* distribution of the burdens of inaccuracy. On this view, inaccuracy is not objectionable if it works to everyone's benefit. Differential burdening of individual insureds, however, may be objectionable. This kind of concern has egalitarian roots, because it is based on the notion that the risk of inaccurate classification should be borne by the community of all insureds, rather than by a few who suffer the entire disadvantage of inaccuracy.[61]

Differential inaccuracy is likely to be the product of coincidental and contingent differences in the ease of obtaining information about different variables that might predict expected loss. It is inexpensive and easy to determine an insured's sex; it is expensive and difficult to determine his or her habits and character. Some variables are thus made the basis of risk classes because they are available and useful rather than because they are more accurate or preferable. The two following examples indicate the different ways

---

rates charging everyone a higher rate to avoid mistreating low risk individuals).

[60] See Posner, A Theory of Negligence, 1 J. Legal Stud. 29 (1972).

[61] This notion might be considered an application of Rawls' "difference principle" to insurance. See J. Rawls, A Theory of Justice 75-83 (1971).

in which these contingencies can affect distribution of the burdens of inaccuracy.

Example 1. Suppose that drunken driving by young males causes a high percentage of all automobile accidents, but only ten percent of all young males drive while drunk. Two risk classes are formed: one for young males and one for all other drivers. Young males pay much higher liability insurance premiums than all other drivers.

FIGURE 4

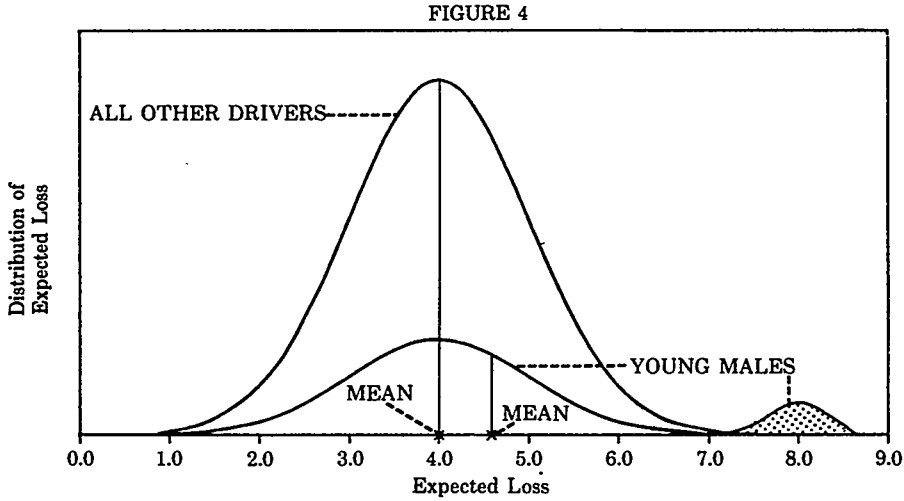


Figure 4 illustrates the two classes. The class of young males in this example is assumed to be much less homogeneous than the class of all other drivers. Ten percent of the former class (represented by the shaded area) has very high expected losses, but ninety percent has average expected losses equal to the average of all other drivers. If there is no economically worthwhile method of discovering which young males have high expected losses, all young males bear the burden of the inaccuracy in this classification scheme. Those young males who have no greater chance of having accidents than other drivers are therefore charged considerably higher premiums. In effect, low-risk young males "subsidize" high-risk young males and all other drivers, simply because they happen to be young and male.

Example 2. Suppose that although the causes of automobile ac-

cidents are uncertain, young males, older males, young females, and older females have different and separable expected losses. Ten percent of each class has substantially higher expected losses than the other members of the class, but it is not feasible to isolate this group in each class. Four risk classes are formed, each paying different premiums.

FIGURE 5

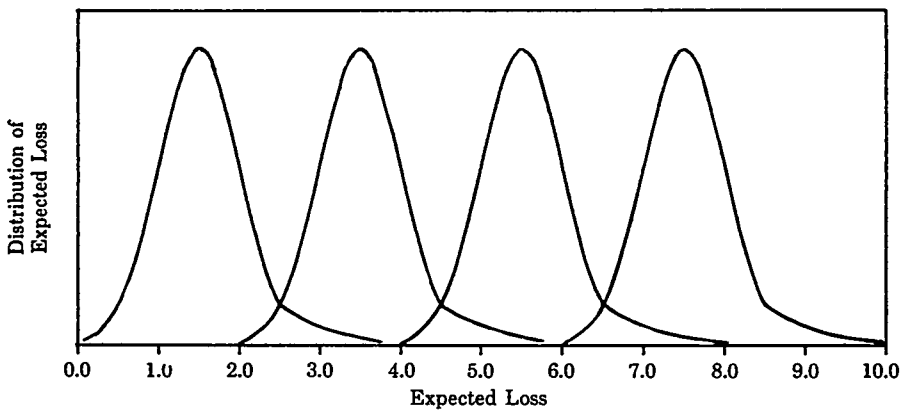


Figure 5 illustrates the four classes. The classes have roughly comparable homogeneity (or heterogeneity, depending on one's perspective). Here, one group does not "subsidize" all the high-risk drivers while all others avoid the burden entirely. Instead, the comparatively low-risk members of each class "subsidize" the few high-risk drivers in their class.

The difference between the two examples is that in the first, a small group of all insureds bears the burden of inaccurate classification. The young males in the first example have a much greater risk-sharing obligation than all others. By contrast, in the second example, all insureds bear a more nearly similar burden of inaccuracy and therefore have similar risk-sharing obligations. In Example 1, comparatively few insureds pay a large additional premium, and all other insureds save a smaller amount.[62] In Example 2, most insureds pay modest additional premiums, and consequently a few insureds pay considerably less than they should. Are the large

[62] Of course, the high-risk young males save a substantial amount.

overcharges of low-risk young males in Example 1 more objectionable than the sizeable undercharges of ten percent of each class in Example 2, or are they equally objectionable?[63]

There is a straightforward "differential inaccuracy" argument against permitting the burden of inaccuracy in Example 1. Allocating a heavy burden to one group of insureds merely because a small percentage of those with similar characteristics has high expected losses fails to treat the members of that group with the concern and respect equal to that accorded other groups. This method of classification uses most members of the disadvantaged group as a means to an end: assuring that the few of their number who are truly high-risk are charged higher premiums. Because considerable inaccuracy is unavoidable in this situation, broader and more even distribution of the risk of being overcharged would more nearly treat all insureds with equal concern and respect. If all insureds were unaware of their own characteristics, this argument goes, enough of them would want protection against the risk of being in the group bearing the burden of inaccuracy so that at least some "tempering" of this burden would be required.[64]

This argument, however, holds only weakly if at all when leveled against the classification illustrated in Example 2. In this example, differential inaccuracy is lower. It is no longer true that only a few insureds are overcharged heavily and are thereby forced to carry the greatest burden of risk sharing. Instead, most insureds are overcharged a modest amount, and the risk of having to bear the burdens of inaccuracy is tempered. More insureds share the risk that some are being inaccurately classified.

Of course, completely even distribution of the burdens of inaccuracy is not feasible without surrendering a great deal of refinement in risk classification. Because the ease of gathering information

---

[63] The insurance industry position, unsurprisingly, is that there is no basis for choosing between alternatives that involve the same aggregate amounts of overcharging or undercharging. The industry's position is essentially utilitarian. The utility or disutility resulting from such inaccuracy, under the industry's view, is simply a function of the total amount of inaccuracy, and not at all related to the way the inaccuracy is distributed. See National Ass'n of Ins. Comm'rs, supra note 19, at 155-74.

[64] This might be called an "egalitarian" classification scheme. For one version of this approach, see Ferreira, Identifying Equitable Insurance Premiums for Risk Classes: An Alternative to the Classical Approach, *in* Massachusetts Div. of Ins., supra note 19, at 74-120. Ferreira would "temper" the differences in homogeneity by adjusting the rates. Id. at 109-10.

about insureds varies according to the kind of information that the system is using, tension between refined classification and equal distribution of the risk of inaccuracy is inevitable. A crucial question, then, is what amount of uneven distribution is tolerable in return for what amount of refinement? Although an answer may not be specifiable in the abstract, it is possible to specify the two separate factors that are relevant.

The first is the amount of homogeneity within the burdened class. The more homogeneous the class, the more tolerable the burden. The nine-to-one ratio of low- to high-risk young males in Example 1, for instance, meant that the class was not very homogeneous. To establish a class with this much variation in expected loss is to identify an entire group by the characteristics of only a few of its members. This kind of classification ignores the fact that ninety percent of the burdened class has an expected loss identical to that of another class that is being charged much less. Nevertheless, standing alone this observation is nothing more than a statement that, other things being equal, more accuracy in general is preferable to less.

The second relevant factor is *comparative* homogeneity. If risk classes are relatively heterogeneous, but equally so, as in Example 2, then the members of one class do not benefit at the expense of the members of a different class. Instead, low-risk insureds in each class bear a similar burden of inaccuracy. If one class is more heterogeneous than others, however, the burden of inaccuracy is spread unevenly. The crux of an accuracy-equity objection to a given classification scheme, then, is the difference in the comparative homogeneity of the risk classes that the scheme has adopted.

It is therefore important to notice that critics of differential risk-sharing are not always voicing accuracy-equity objections. Differences in the comparative homogeneity of risk classes do not necessarily underlie such attacks. For example, sex-based life insurance and pension classifications often are attacked because about eighty-five percent of men and women have the same life expectancies.[65] Because the homogeneity of the two classes is roughly similar, however, an accuracy-equity concern cannot be the basis

[65] See, e.g., Brilmayer, Heckler, Laycock & Sullivan, supra note 2, at 531. The degree of overlap is variously stated as being between 80% and 84%. See Kimball, supra note 2, at 121 & n.100.

for these attacks. If the only reason for the difference between male and female life expectancies were that fifteen percent of all females lived longer than all other females and all males, then sex-based classification would clearly raise accuracy-equity concerns. But this greater female longevity is not the only reason for the difference in life expectancies. The separation between the classes occurs not only at the above-average life expectancy end of the spectrum, but also at the below-average end. That is, the difference between male and female life expectancies is caused not only by a minority of females who have higher than average expectancies, but also by a minority of males who have lower than average expectancies.[66] The two classes actually have roughly comparable homogeneity.

Both long-expectancy males and long-expectancy females consequently "subsidize" the other members of their respective classes when they purchase life insurance. Both short-expectancy males and short-expectancy females "subsidize" the other members of their class when they purchase pensions. Thus, if sex-based classification in life insurance and pensions is to be soundly criticized for its unfairness, that criticism must be based on the grounds to be explored in the next two sections.

B. Control-Causality

The second relevant issue regarding the distribution of risk through insurance classification is whether variables that neither are within the insured's control nor are causes of loss should be used to form risk classes. Noncausal variables can be criticized because of their apparent arbitrariness; they appear merely to be statistically correlated with expected loss.[67] Sometimes, control and causality are conceptually distinct notions, and the major focus of criticism has to do with controllability. On other occasions, however, the two notions are harder to distinguish, and reference to "control-causality" objections is more accurate.

---

[66] For a table of life expectancies, see Bureau of the Census, U.S. Dep't of Commerce, Statistical Abstract of the United States 73 (104th ed. 1984).

[67] See generally Underwood, Law and the Crystal Ball: Predicting Behavior with Statistical Inference and Individualized Judgment, 88 Yale L.J. 1408 (1979) (advocating distribution of benefits and burdens based on predictions of individual behavior only when predictions rest on controllable factors, due to respect for personal autonomy).

Noncontrollable variables can be criticized on the ground that their use makes the exercise of individual responsibility irrelevant to the price of insurance.[68] No amount of care or safety and no period of loss-free behavior can affect the cost of coverage when such variables distinguish risk classes. Thus, the use of noncontrollable variables denies individuals the opportunity, through the exercise of individual responsibility, to alter the effect of being "grouped."

The choice as to which variables raise these control-causality problems consequently will depend on the critic's conception of individual responsibility and controllability. In common parlance, people are not "responsible" for their gender, but they are "responsible" for and can control their smoking and eating habits. In between these two extremes, lie a variety of actions that are more difficult to characterize. Some people would argue that individuals are responsible for any characteristics that are even theoretically subject to change—only such characteristics as sex, age, race, and birth defects are considered beyond one's control. In contrast, a more egalitarian conception of responsibility might also consider the results of the "natural lottery" in skills and talents to be beyond individual control.[69] Under this view, even variables bearing a causal connection to insured losses would be unacceptable if adequate "controllability" were missing. A driver's congenitally slow "reaction time" may be a cause of accidents, but it is not within his control. Only the difficulty of distinguishing slow-reactioned from quick but careless drivers could then justify penalizing the former for errors caused by their slow reactions.

One might argue that even if a particular variable is not under

---

[68] A different criticism of the use of such variables is that they may deny insureds their just deserts, because they treat insureds as members of groups, rather than as separate individuals with autonomy and personality. The weakness of this criticism is that an objection to treating individual insureds as members of groups is not a criticism of any particular form of classification, but to classification itself. To be sure, any system that uses a specified set of variables can be altered to take additional variables into account. Such a move, however, will not constitute treating insureds as individuals. Instead, it will treat insureds as members of yet another set of groups—those whose members do or do not have the characteristics specified by the additional variables. If automobile classifications take only age into account, or only age and sex, then the addition of territory of residence and mileage driven annually will respond to the criticism that relevant characteristics were being ignored. These additions, however, will not result in the treatment of insureds as individuals.

[69] For discussion of the "natural lottery" see J. Rawls, supra note 61, at 72 (discussing the distribution of assets in a system of natural liberty).

an insured's control, his activity level is, and that controllability therefore is never a concern. The same kinds of considerations, however, operate here as well. Although whether to drive at all is theoretically a matter of choice, driving is so often a necessity that having at least one car is probably not controllable in any practical sense. The element of choice present in most acts of responsibility or irresponsibility is almost entirely absent from this one. The number of miles driven or the number of cars owned by a family is more clearly determined, however, by free choice.

In short, the notion of "control" is a normative conclusion about the characteristics for which individuals can properly be asked to bear insurance responsibility. This notion does not solve risk-distributional issues; rather, it directs one's attention to them. Do poor people "control" where they live? Do women "control" where they work or how much they drive? People with radically different views about the answers to these questions will be unable to agree on what constitutes a fair or acceptable classification variable.[70] Conducting the debate about risk-distributional fairness on these terms is likely to be more fruitful, however, than reliance on notions like "discrimination" or "statistical justification." These notions normally are not sharp enough—in this context at least—to help isolate what is at stake in the controversy over classification.[71]

The first step out of the quandary over how to define the notion of individual responsibility may be to recognize a few *ceteris paribus* conditions that are acceptable to many different points of view. Although these conditions do not resolve the tensions between different conceptions of responsibility, they do provide a method of ranking alternatives in terms of preferability. The first condition is that the more random the losses in question (in the sense defined earlier),[72] the less reason there is to object to any particular classification variable. When such variables account for only a relatively small portion of actual losses, the effect on premium differences they produce will not be great. For example, the driver's sex accounts for over one-third of the explainable portion

---

[70] For a discussion from a critical legal studies viewpoint, see Austin, supra note 3.

[71] Although I disagree with Professor Austin's conclusion that these problems are unresolvable, see generally id., in this sense I agree with her characterization of the issue at stake in classification controversies.

[72] See supra p. 409.

of "variance of expected loss distribution" among drivers—a statistical measure of the efficiency of the variable.[73] Because all variables in use account for only twenty-two percent of that variance,[74] however, the vast majority of accidents must be treated as occurring at random.[75] One study showed that if sex were eliminated as a variable, young female drivers' automobile insurance rates would increase twenty-six percent and young male drivers' rates would decrease six percent (females are only twenty-four percent of the youthful driving population).[76]

Elimination of sex as a driver classification variable because of its noncontrollability may well thus satisfy many control-causality objections. It should be noted, however, that a more efficient variable cannot necessarily be substituted whenever a noncontrollable variable, such as sex, is eliminated. Often a classification system uses an objectionable variable precisely because it is efficient. This point is ignored too often by those who demand that a classification system count past loss experience more heavily in setting premiums. Experience rating cannot safely exceed the credibility of the loss experience on which it is based without risking gross inaccuracy. When the average insured suffers a loss only once every ten or twelve years—as is the case in automobile liability—then the occurrence of a loss only marginally increases his expected loss.[77] The probability is still substantial that the loss was the product of chance. On the other hand, in fields where losses occur more frequently, experience rating becomes more feasible. Because this form of classification leaves the insured with maximum flexibility to affect premium levels through the exercise of individual responsibility, control-causality objections should diminish as reliance on experience rating increases.

A second *ceteris paribus* condition is that a shift from a noncontrollable to a controllable variable becomes more preferable the

---

[73] National Ass'n of Ins. Comm'rs, supra note 19, at 142. The variance of expected loss distribution is a measure of the differences between the insureds' individual expected losses and the group's average expected loss.

[74] Id.

[75] See id. at 177 (emphasizing the value of territorial ratings, which account for 7% of expected loss variance, or almost one-third of the 22% total).

[76] See id. at 26.

[77] See supra note 48 and accompanying text; see also National Ass'n of Ins. Comm'rs, supra note 19, at 61-66 (showing inadequacy of driving experience "merit rating" as a substitute for the standard age/sex/marital status classifications).

less it hinders efficiency. The principle underlying this rule may at first glance seem counterintuitive because a move to a more controllable variable suggests greater, not less efficiency. It is only superficially a paradox, however, that a controllable variable could produce less efficiency. A noncontrollable variable, such as the insured's sex, may be a convenient surrogate for controllable variables, such as safety in one's workplace or the number of miles one is likely to drive.[78] But controllable variables, such as the latter two, may be expensive to use because of their unreliability and thus less efficient than their cruder and otherwise more objectionable surrogates. On the other hand, when the advantage of using noncontrollable variables is slight, a shift to controllable variables will sacrifice little efficiency and will increase the perceived fairness and legitimacy of the classification system. Yet it is unlikely that any insurance company would make such a move on its own if the move were even marginally inefficient. Legal compulsion would be required.

It follows that some noncontrollable variables might be prohibited with relatively little loss of efficiency, whereas the elimination of others could be more costly. The amount of the loss would depend on the degree to which the variable in question is a surrogate for a controllable variable, and on the cost and reliability of assigning insureds to classes based on a substituted variable. For example, it is often assumed that the prohibition of sex-based life insurance or pension classification will automatically result in the abolition of all classifications other than age. Males would then subsidize female pension rates, and females would subsidize male life insurance rates.[79] This result will occur, however, only if sex stands for no other readily discoverable and easily substituted variable or variables. If instead, sex is a surrogate (for example) for smoking habits or exposure to hazardous substances in the workplace, then after its prohibition as a variable, other variables may eventually replace it. Many insurance companies already take such variables into account in setting rates.

A third and related condition is derived from the fact that noncontrollable variables do not preclude the exercise of individual re-

---

[78] See, e.g., National Ass'n of Ins. Comm'rs, supra note 19, at 321 (alcohol consumption associated with age).

[79] See Kimball, supra note 2, at 119-23.

sponsibility for the control of losses; they merely make the exercise of such control irrelevant to the cost of any given individual's insurance. Insureds whose rates are determined by noncontrollable variables have an incentive to shift their protection dollars away from insurance and into less-expensive loss prevention.[80] The greater the marginal rate of this substitution, the stronger the inference that banning the noncontrollable variable would be feasible, because the fact of substitution is evidence that a readily controllable variable exists. The lower the rate of this substitution, the stronger the inference that the costs of loss prevention are too high to warrant the substitution of loss prevention for insurance.[81] In this situation, control-causality objections to noncontrollable variables should diminish, because there are no generally suitable replacements for those in use.

C. Suspect Variables

The preceding sections explored two separate reasons why otherwise useful classification variables may be considered inadmissible. Even when these accuracy-equity or control-causality objections are weak or nonexistent, however, certain variables are likely to remain under attack. Symbolic and principled rather than consequentialist concerns often cause criticism of these "suspect" variables. It is easiest to see the independence of this type of objection from the other two types by considering an extreme case. For example, even if classification based on race raised neither of the other two risk-distributional concerns, it would be unacceptable to most people on symbolic and principled grounds.

Whenever an efficiency-promoting but emotionally or morally suspect variable is rendered inadmissible, a redistribution of risk occurs. The risk assessment goal of insurance is sacrificed to further a risk-distributional goal. Subsidies then run from certain low-risk insureds to certain high-risk insureds. Variables that would otherwise be used in classification because of their economic superiority are not used because of their social inferiority. The efficiency lost is simply the cost of avoiding "discrimination" in insur-

---

[80] See supra p. 425.

[81] For evidence of this rate of suhstitution, one would have to look at the changes in coverage levels purchased by individuals after institution of new classification schemes. Because several states have adopted new schemes, such data may eventually be available.

ance on the basis of the suspect variable. Indeed, our willingness to ignore variables that would otherwise be useful for classification is largely what defines them as discriminatory in this context. For example, if we had decided that men and women should be treated equally as *groups*, then sex-based classification would not only be acceptable, but might actually be required.[82] When treatment of men and women as *individuals* is the predominant goal, however, such classification is likely to appear "discriminatory."

Classification variables may be suspect for a variety of reasons. First, a particular characteristic may be used improperly in other fields and therefore be objectionable on symbolic grounds. For example, because women often have been subjected to discriminatory treatment in other social and economic settings, use of sex as a classification variable may be objectionable. Second, the data on which predictions of expected loss are based may not have enough probative power to justify their use in drawing socially questionable distinctions. Life expectancy calculations for men and women may be based on mortality rates formulated before recent improvements in health and safety features of male-populated workplaces. This concern can also be viewed as an "efficiency" concern, but doing so does not vitiate the moral suspiciousness of the classification. Third, some variables may be suspect because they are used only to the disadvantage of certain groups, but never to their advantage. For example, women may have been charged more than men for pensions, but the same for life insurance. Finally, certain variables may help to perpetuate unfair disadvantages operating outside the insurance system. If blacks have lower life expectancies than whites, the reason may be that they have less healthful living and working conditions, due perhaps to racism and bigotry. If race-based variables were not already inadmissible for other reasons, this factor would render them suspect.

Nevertheless, declaring a variable inadmissible does not necessarily eliminate all the effects that had been associated with its use. The symbolic effect of using a variable is eliminated upon its pro-

---

[82] Compare Rutherglen, supra note 2 (distinguishing disparate treatment and disparate impact, and concluding that equal treatment of individuals is the proper guideline) with Brilmayer, Hekeler, Laycock & Sullivan, supra note 2 (legislatures should be free to choose between anticlassification and the recognition of sexual differences to achieve equal treatment; *Manhart* represents a correct result due to practical considerations).

hibition. If the concern underlying a variable's inadmissibility is its actual disadvantaging effect, however, then the choice of a variable to replace it is critical. For example, when territory of residence functions as a close surrogate for race in automobile insurance, differential treatment on the basis of race has been symbolically eliminated, but for many persons the disparate treatment continues in fact through the use of territorial variables. Whether the use of territorial variables is considered an acceptable alternative to discriminatory classification or a mere subterfuge depends on whether the purpose of the prohibition was symbolic or practical. Furthermore, even when the variables in use are not surrogates for others that are inadmissible, marketing practices can circumvent a mandated classification scheme, if suspect variables are easily recognized.[83] Racial or sexual bias in the marketing of coverage is not automatically foreclosed when a classification system is purged of suspect variables.

Because the degree to which a variable is suspect depends on both the nature of the variable and the insurance setting in which it is employed, determining whether a variable should be admissible often poses vexing problems. The use of sex as a classification variable is a good example. My view is that the insured's sex should be inadmissible under almost all circumstances, but this position cannot be ascribed to a single consideration.

In the life insurance and pension fields, as we saw earlier, there are probably no substantial accuracy-equity objections to sex-based classification.[84] There are, however, significant control-causality objections to sex-based classification for life expectancy purposes, although I have not seen the efficiency of alternative controllable variables persuasively demonstrated. As a result, objection to sex-based classification in these lines of insurance must rest heavily on the intensity of one's principled objections to this form of classification as suspect in and of itself.

In automobile insurance, by contrast, there probably are significant accuracy-equity objections to the manner in which young males are sometimes classified.[85] In addition, supplementing sex-

---

[83] See supra note 53.

[84] See supra text accompanying note 64.

[85] In other words, the treatment of young males more closely resembles Example 1 in Part IV.A.2 of this Article than Example 2. See supra pp. 432-33; Ferreira, supra note 64, at 75-77.

based classifications with more controllable variables or greater experience-rating might at least temper objections to sex-based classification. On the other hand, males rather than females are disadvantaged by sex-based automobile classification. To the extent that objections to such a possibly suspect classification arise because of past discrimination against women, they should be weaker here than in connection with pension classification, where women bear the disadvantage of sex-based classification.

A different consideration is that greater inefficiency in automobile insurance seems likely to result in at least marginal increases in personal injuries, whereas the principal effect of any inefficiency in life insurance or pensions will merely be a change in investment patterns. Thus, different considerations may come into play in different insurance contexts. Nevertheless, the symbolic effect of eliminating sex-based classifications only when they disadvantage women would probably be intolerable. If some such classifications are eliminated, all should therefore be abolished.

Notwithstanding the generality of this conclusion, it is worth recognizing that whether a variable is suspect may depend for some purposes on the direction in which the benefits and burdens it creates run.[86] This recognition highlights a separate and final risk-distributional ground for analyzing a classification system: the use of classification for the express purpose of creating benefits for certain groups. Notice that sacrificing efficiency in classification for this purpose is not simply a means of assuring risk-distributional fairness. Rather, such moves constitute an affirmative redistribution of risk, and occupy our final category.

D. Risk-Redistributional Policies

Insurance classification can distribute risk not only by ignoring or making inadmissible certain variables that are linked with expected loss—such as race, sex, age, and marital status—it can also further a variety of risk-redistributional goals. Classes can be constructed and prices set so that some classes are not self-supporting and others are overcharged. Cross-subsidies then run from the overcharged to the undercharged classes. The old can be made to

---

[86] See supra note 37 (age discrimination may be acceptable when it benefits rather than burdens the elderly).

subsidize the young (or vice versa), the healthy the sick, or careful insureds the careless. Moreover, if one class is largely a surrogate for another recognizable group—for example, when territory of residence substitutes fairly accurately for race—then racial groups may be over- or undercharged without explicit reliance on a suspect variable. If a good enough surrogate for wealth can be made the basis of risk classes, the rich may even be separately classified so that they subsidize the poor, or the poor the rich. In addition, one line of insurance can be used to subsidize another—products liability rates may be set high and auto liability rates low across-the-board.

This form of risk redistribution is not designed to avoid differential treatment of certain risk classes. On the contrary, it is intended to accord special treatment to the classes that benefit to increase their welfare. In so doing, such classification may raise accuracy-equity, control-causality, or suspect variable problems of its own. Risk-redistributional classification would create such problems deliberately, however, rather than as a byproduct of the search for efficiency. This approach seeks to benefit certain persons or groups; inevitably burdens must be placed on others if this goal is to be achieved. Any attempt to modify an economically efficient classification scheme to redistribute risk or achieve other goals of public policy, however, is likely to encounter considerable obstacles. We now consider these obstacles.

V. The Costs of Risk-Distributional Fairness

Any society that relies on the private market to supply insurance faces real difficulties when it attempts to assure fair classification or to promote other goals of public policy through legal regulation of classification practices. The reason is that the economic interests of insurers and potential insureds do not simply disappear after a reform is adopted. Instead, these interests continue to operate and easily may confound the objectives of the system. We have already seen such forces at work. An individual insurer has only a limited incentive to discover new methods of classification. The cost of discovering new expected loss data may not warrant the advantages to be gained from the discovery, and other insurers can free-ride on the innovating insurer's investment in new forms of

classification.[87] The solution to this problem of externalized benefit does not always lie in declaring existing classification variables inadmissible in the hope of forcing insurers to develop replacement variables on a now-clean slate. A better solution is collective investment in the discovery of new forms of classification that are more efficient than those now in use. Where private collective enterprises, such as insurance trade associations, have not developed new classifications, then governmental investment in such development might be worthwhile.

A second problem is that any form of limited or partial regulation, such as redistributional classification, is susceptible to circumvention by market forces. Sometimes, the form taken by market circumvention of regulatory constraints can be predicted. If the advantages of such market adjustments outweigh their disadvantages, regulation may be deliberately designed to encourage these very adjustments. More often, however, remedying concerns about accuracy-equity, control-causality, or suspect variables, or implementing wholesale risk redistribution will require swimming against strong economic currents. Consider the efforts that would be necessary.

First, only a universally adopted scheme could endure. Any insurer who voluntarily adopted a classification system that was noticeably counter-efficient would quickly find itself afflicted by adverse selection.[88] Low-risk insureds who were offered less expensive coverage by other insurers would depart, and high-risk insureds would gravitate in its direction. The greater the cross-subsidies, the more likely it is that a voluntary scheme ultimately would ruin the insurer adopting the scheme.

Second, even if a counter-efficient classification scheme were mandatory, all insurers would be subject to a different adverse selection effect. Recall that insurers compete not only among themselves for premium dollars. They also compete with noninsurance methods of loss protection for available protection dollars. A mandatory redistributional scheme will encourage insureds in the subsidizing classes to shift their protection dollars away from insurance and into other methods of loss protection. An insured who

---

[87] See supra text accompanying note 44.

[88] Cf. supra note 17 (insurer who fails to copy an efficient innovation will suffer adverse selection at hands of high-risk insureds).

had previously maintained $10,000 of coverage may now find that his dollars are better spent by purchasing only $8,000 of coverage and investing the money saved on safety precautions that promise more than an additional $2,000 of expected loss protection.

The effects of this option to bail out of the reformed system would depend on the cross-elasticity of demand for insurance and loss prevention.[89] The more easily low-risk insureds could substitute additional investment in loss prevention for insurance coverage, the greater the threat that these substitutions would undermine the system by reducing the fund that finances the subsidies. Correspondingly, insureds in subsidized risk classes would be encouraged to invest less in prevention and more in insurance than they had previously. As a result, there would be not only more losses, but more losses covered by insurance.[90] Even higher rates and possibly greater subsidies might then be required.

This system could eventually fall under its own weight, or reach an equilibrium in which a few subsidizers buy small amounts of coverage and many modestly subsidized insureds buy a little more coverage than they would otherwise. Regardless, there would be much less cross-subsidization than the modified system originally had envisioned, and members of burdened risk classes would purchase less insurance than would otherwise be optimal.

Furthermore, unless there were also some constraints on the discretion of insurers to accept or reject applicants for coverage, underwriting decisions would tend to neutralize the effects of the classification system.[91] Applicants eligible for placement in the subsidized classes would be rejected with disproportionate frequency. They would find it generally difficult to purchase the subsidized coverage for which the rules of the system made them eligible. Only strict enforcement of mandatory offer and acceptance rules could restrict this phenomenon.

In short, the redistributional classification scheme described thus far would have holes through which those supposedly burdened with subsidy obligations could escape. Another device would

---

[89] For a discussion of the concept of cross-elasticity, see R. Posner, Economic Analysis of Law 222 (2d ed. 1977).

[90] Increased losses would be caused by decreases in safety, increases in activity levels, or both.

[91] See National Ass'n of Ins. Comm'rs, supra note 19, at 16.

therefore need to be introduced to stabilize the system. Not only would participation in a redistributional classification have to be mandatory for insurers, but the purchase of insurance would also have to be required. Yet such a requirement is realistic in only a few fields.

When the avenues of escape from the system can be closed by making the purchase of coverage mandatory, a number of different redistributional mechanisms can be adopted.[92] One is an "assigned risk" system under which individuals falling into a subsidized category are assigned in rotation to different insurers, which are then required to insure them at the specified subsidy rate. Rates charged other classes are correspondingly higher than they would be in the absence of the subsidy. A different approach uses a governmentally-operated fund that is financed by surcharges against insurers and which provides coverage to group members who are subsidized. Rates on the private market are again correspondingly higher because of the surcharges that finance the residual-market subsidies. Finally, a hybrid of the first two approaches is a "joint underwriting association," financed and operated by all the insurers of the jurisdiction. The association writes coverage at subsidized rates for specified classes of insureds. Under any of these schemes, low-risk insureds tend to purchase less coverage than in the absence of the subsidies and tend to spend more on other forms of protection, which appear comparatively cheaper than before.

This analysis suggests the numerous difficulties entailed in remedying risk-distributional unfairness or in producing significant amounts of risk redistribution by modifying risk classification practices alone. Voluntary schemes are likely to fall far short of their goal or, even worse, to bankrupt their participants. Programs made mandatory for insurers may have considerable effect, but only at the cost of reducing the amount of insurance coverage that is purchased—in the opinion of many people, an undesirable development. Remedies promoting significant risk redistribution or firmly enforcing risk-distributional fairness may have to make insurance mandatory not only for insurers but also for insureds. In the end this last point is intuitively obvious, because risk redistri-

---

[92] For a catalogue and analysis of these devices, see U.S. Dept. of Commerce, supra note 32.

bution is a kind of tax scheme, and when a voluntary activity is taxed, it is likely to draw fewer participants than substitute activities that are not subject to the tax.

This analysis has three related implications. First, remedying accuracy-equity, control-causality, or suspect variable concerns often will be feasible, but only at a cost. We should recognize this cost, rather than assume that the only effect of remedying unfairness is to equalize burdens or benefits. A certain amount of inefficiency in the entire system, a corresponding loss in safety, or an increase in the level of unsafe activities may be part of the price that must be paid to achieve a fairer distribution of risk.

Second, sometimes the risk in question is one that is appropriately redistributed only within the insurance system. On other occasions, however, the risk is the kind that society as a whole should bear. Accuracy-equity concerns normally fall into the former category because they arise mainly from the difficulty of determining expected loss. Some control-causality and suspect variable concerns, on the other hand, arise from commonly-held notions of equality the costs of which should accordingly be distributed broadly. If the form of insurance involved is nearly universal—mandatory automobile coverage, for example—perhaps the risk-sharing population is already broad enough to warrant using the insurance system as a partial remedy for the inequality in question. Because so many persons participate in that kind of coverage, in effect society as a whole is bearing the costs of the redistribution. But if the insurance involved is specialized—professional liability or private disability coverage, for instance—then it is not at all clear that the cost of redistributing risk within the system should be borne only by those participating.

An example may help illustrate this point. Suppose we decided that hemophiliacs and nonhemophiliacs should be charged the same rates for disability insurance. In such a case there would be no particular reason why disability insureds alone should be forced to share the risk that we have decided hemophiliacs alone should not have to bear. The decision to protect hemophiliacs is designed to remedy not a risk-distributional unfairness inherent in disability insurance, but an unfortunate outcome of the "natural lottery" for physical health. Unless there already has been an across-the-board effort to subsidize hemophiliacs, of which the change in disability insurance classification is but a part, it seems inappropriate to ask

disability insureds to bear the full cost of subsidies to hemophiliacs. Where more generalized efforts at redistribution of a particular risk, however, have occurred—by eliminating sexual or racial discrimination at large, for instance—asking the holders of specialized forms of insurance coverage to bear their share of the cost of this effort is much less problematic.

This insight leads to a final point. There are much simpler ways of redistributing risk wholesale, if that is desired, than through regulation of insurance risk classification. Members of risk classes deserving subsidies can simply be given money out of selectively assessed taxes and allowed to use it as they wish. If some forms of insurance are considered "merit" goods—those that society as a whole believes people should have regardless of whether they would choose to purchase them—then insurance can be bought for the classes whom we wish to subsidize. For example, certain groups might be issued "insurance stamps" that would purchase insurance only. As an alternative, the government could finance an insurance system to cover those whom it considers desirable to subsidize. If certain forms of insurance are not merit goods, but subsidies to some groups are nevertheless considered desirable—perhaps because the market improperly overcharges them—governmentally provided reinsurance may be a workable solution.

These more straightforward methods of redistributing risk are likely to be much more effective at wholesale redistribution than is a method that relies on risk classification to do the same job. More direct forms of redistribution that do not rely on insurance at all, such as flood control, may be even more effective in particular situations. These methods, however, are not always politically feasible. The legal system has thus developed other devices for achieving small-scale, or "local" redistribution. Making regulatory adjustments in the risk classification system is one such method. These ad hoc methods of risk distribution may be inefficient, subject to circumvention by the market, and only partially effective. They have the virtue, however, at least for proponents of risk-redistribution, of being possible to implement. More than anything else, this may explain their attractiveness.

VI. Conclusion

Evaluating the efficiency and fairness of insurance classification systems is a complex undertaking. Different notions of fairness are

likely to lead to different evaluations, and the claims of each notion of risk-distributional fairness must make their peace with the inherent demands of any classification system and with countervailing considerations of economic efficiency. Although efficiency may be a more monolithic concept in this context than risk-distributional fairness, the safety incentives produced even by efficient classification may be less powerful than desired. Limitations on available information and the diminishing returns of increasingly refined classification make pricing in accordance with expected cost an ideal, but never a wholly attainable one.

Because a classification system is unavoidably influenced by the effects of market incentives, attempts to make it more fair or to redistribute risk through regulation of risk classification and pricing require significant legal interventions to be successful. Mandatory insurance and government operation of substitute or residual insurance markets may even be necessary to maintain an effective redistributional scheme. In the end, it is therefore likely that the compromise between efficiency and the broad distribution of risk that is inevitable in any insurance system will never fully satisfy the proponents of either value.