The Liability Crisis and the Dynamics of Competitive Insurance Markets

Ralph A. Winter†

From late 1984 to mid-1986, the liability insurance industry was characterized by dramatic changes in the pricing and availability of insurance coverage. Premiums in commercial casualty lines of insurance rose by hundreds of percent for some policies. Limits on coverage in many policies were reduced to a fraction of potential losses and insurers stopped covering some types of risks and entire lines altogether.1

Although the insurance crisis of 1984–86 was the most severe episode of rapidly rising premiums and decreasing availability, it was not the first. The industry has a pattern of alternating between soft markets characterized by stable premiums and low rates of return to insurers, and periodic tight markets with high premiums and returns. From about 1977 until 1985, the industry was in a soft market with declining returns, which was preceded by a tight market in the mid-1970s, which in turn was preceded by a soft market since the late 1960s.2

The liability insurance crisis and the general recognition of an insurance cycle have been met with developments on two policy fronts, tort law reform and insurance rate regulation. More than forty states have enacted tort law reforms in the last two years in attempts to change the basic market conditions that most observers believe are at the heart of the insurance crisis. In 1986 New York moved to “flex-rate” insurance regulation, which requires sixty days notice for premium increases of more than ten percent, and California, a state with traditionally less regulation, has pro-

† Professor of Economics and Finance, University of Toronto and Olin Fellow in the Civil Liability Program, Yale Law School. Support for this work from an Olin Fellowship in the Civil Liability Program at Yale Law School is gratefully acknowledged. I am grateful as well for comments from Scott Harington, Yehuda Kotowitz, Jeff Perloff, George Priest, Tom Scott, and Don Whitman. Jane Gross, Bill Biggs, Lisa Olivetti, Bill Barnes, and Rachel Siegel did a superb job of editing this Article. None of these individuals is responsible for remaining shortcomings.


2. The cyclical pattern of insurance availability, premiums and returns is described in Stewart, Profit Cycles in Property-Liability Insurance, in 1 ISSUES IN INSURANCE 273, 303-04 (J. Long & E. Randall 3d ed. 1984); see infra note 33, REPORT OF THE GOVERNOR’S ADVISORY COMM’N ON LIAB. INS., INSURING OUR FUTURE 70-75 (1986) (State of New York) [hereinafter CUOMO COMM’N].
posed similar policies. As of March 1987, thirty-seven states had placed restrictions on the cancellation or nonrenewal of liability insurance policies.8

The apparent goal of these regulations is to mitigate sudden premium increases and availability problems. Parallel to these state level developments is a federal level movement to repeal the McCarran-Ferguson Act's4 exemption of the insurance industry from federal antitrust statutes.5

The proposals to change government policy towards the insurance industry are based on an assumption that the crisis is the result of market failure in the industry, rather than the response of a competitive industry to changing market conditions. This position is consistent with the popular view that the insurance industry fluctuates between competitive and collusive pricing or between price wars and adequate pricing.6

This Article offers an economic analysis of the dynamics of premiums and availability in the liability insurance market and considers the implications of this analysis for public policy. The feature of the liability insurance market that is critical for explaining the pattern of periodic tight markets is the substantial uncertainty that insurers face in predicting claims. In an insurance market in which average claims are uncertain, the net worth of stock insurers measures their capacity to write insurance.7 Crises such as that of the mid-1980s are triggered by a depletion of this capacity because of cumulative losses. Capacity will not adjust immediately in a competitive insurance market because of the cost advantage of retained earnings or internal capital over the issuance of outside equity in raising the net worth of firms in the market. Competitive, profit-maximizing insurers will wait for the rapid accumulation of retained earnings during a tight market with high premiums and profits, rather than resort to issuing enough outside equity to eliminate the capacity constraint.8 While financial capital in the insurance market is fungible compared with

6. Versions of this theory abound in the popular press. For example, Business Week states: Even while the industry [in 1985 and 1986] was blaming its troubles on the tort system, many experts pointed out that its problems were largely self-made. In previous years the industry had slashed prices competitively to the point that it incurred enormous losses. That, rather than excessive jury awards, explained most of the industry's financial difficulties. Now that insurers have been able to push up prices, profitability is returning. Nothing complicated about that.
7. See infra text accompanying note 63.
8. See infra Part III, section B.
capacity (physical capital) in other product markets, a theme of this Article is that even modest rigidities in this capital allow dramatic swings in pricing and availability in response to shocks to capacity.

Explanations of the recent insurance crisis that are based on capacity—that the crisis was the effect of cumulative losses in the market in the early 1980s—have been rejected by substantial law and economics literature on the basis that insurance premiums should reflect future claims, not past losses. This Article argues that the capacity constraint hypothesis is necessary to explain the empirical features of the crisis. In particular, the increased profits during the crisis, the increased issuance of equity by insurers in the capital markets in 1985, and the suddenness of the insurance crisis relative to trends in underlying tort claims cannot be explained without a capacity argument.

The reconciliation of the increased profits of the industry over the period of the crisis and the larger magnitude of premium increases as compared to insurers’ costs is critical for the assessment of government policy; the regulatory response to the crisis has been strongly influenced by the public perception that the premium increases were “unjustified.”

This Article shares with the existing literature on the crisis the conclusion that the crisis was not the consequence of an insurance conspiracy. Its theme is that fluctuations in premiums and availability of insurance is inevitable in an environment of uncertainty such as that resulting from the unpredictability of common tort law. Even the periodic jumps in premiums by more than expected claims (as evidenced by increased profits) are consistent with the competitive structure of the liability insurance market.

Furthermore, the cyclical nature of the market is not a basis for regulating insurance rates. Regulation that prevents premiums from rising to competitive levels has a detrimental impact on availability and on future premiums, which more than offsets any benefit of lower current rates.


   Insurers, like all profit maximizing companies, charge the price which maximizes their profits. Past gains or past losses are irrelevant to setting the price today which will maximize profits tomorrow. The argument that insurers are charging higher premiums to recoup past losses suggests that absent such losses their premiums would be lower—that is, that they would not be charging premiums that maximize their profits. That makes little sense.

10. This Article does not argue that the capacity constraint hypothesis is sufficient to account for the entire magnitude of the crisis of 1984–86. Without question, tort awards rose from 1984 to 1985. See Tort Policy Update, supra note 1, charts I and J. Thus, the size of expected claims increased in the market. But an increase in expected losses in an insurance market raises the demand price for insurance as much as the supply price; therefore it cannot account for other aspects of the crisis such as the withdrawal of insurers from some lines or the tighter limits on coverage.

The onus of improving the performance of the liability insurance markets lies with stabilizing the environment of the market through tort reform, not insurance regulation.

Part I of this Article reviews the empirical dimensions of the recent insurance crisis and of the insurance cycle in general. It then assesses the existing explanations of the crisis in light of the empirical evidence and offers the conclusion that these explanations do not fully account for the empirical features of the crisis. Part II offers a model of the dynamics of a competitive insurance market. Part III then summarizes the specific results of the positive theory outlined in this Article and discusses its implications for four types of public policy: solvency regulation, government policy to increase insurance industry capacity, premium rate regulation, and repeal of the McCarran-Ferguson Act.

I. The Insurance Crisis: Empirical Dimensions

The property-casualty insurance industry had revenues of 176 billion dollars in 1986. Of this amount, forty-one percent was in automobile insurance (1987) and about thirteen percent was in the lines of medical malpractice and general liability insurance. These latter are the lines in which the crisis took place.

A complete explanation of the performance of the liability insurance market in 1984-86 must account for the following eight aspects of the behavior of market prices and availability: (1) the increase in premiums; (2) the drop in insurance transactions for some risks; (3) the tighter limits on coverage; (4) the higher industry profits; (5) the suddenness of the crisis, following a period of stable premiums; (6) the restriction of the crisis to particular lines; (7) the occurrence of the crisis in Canada as well as the United States; and (8) the recurrency of insurance crises, with the industry alternating between crises of high premiums/returns and periods of low returns.

The first aspect that must be accounted for is the increase in premiums. Between 1984 and the end of 1985 premiums increased dramatically in commercial casualty lines of insurance. The increase in premiums is

13. Id. at 103-4.
14. Id. at 103.
15. See Angoff, supra note 5, at 399-400.
17. For instance, municipalities faced premium increases of 100% to 400%, local transit systems faced increases of 500% to 1000%. Medical specialists such as obstetricians and nurse-midwives faced similar increases when they were able to get insurance. Professionals such as engineers and architects
reflected in the increase in revenue (net premiums written) in the Commercial General Liability insurance line shown in Table 1. The revenue for this line rose by seventy-one percent between 1984 and 1985.  

<table>
<thead>
<tr>
<th>Year</th>
<th>Net Premiums Written (Billions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1981</td>
<td>$ 6.0</td>
</tr>
<tr>
<td>1982</td>
<td>5.6</td>
</tr>
<tr>
<td>1983</td>
<td>5.7</td>
</tr>
<tr>
<td>1984</td>
<td>6.5</td>
</tr>
<tr>
<td>1985</td>
<td>11.1</td>
</tr>
</tbody>
</table>

This increase in revenue reflects the increase in premiums in the market, together with the low short-run elasticity of demand in the market. In particular, higher revenue cannot be attributed to an increase in demand and quantity of insurance transacted in the market since the quantity of coverage almost surely fell in 1985 as insurers imposed severe coverage limits and pulled out of many risks entirely that year.

The second aspect is that of the availability of insurance and the drop in insurance transactions for some risks. In lines of insurance where the crisis was most severe, insurance transactions fell nearly to zero. Many municipalities were unable to obtain insurance at prices they were willing to pay; about forty percent of day-care centers had their policies cancelled; the insurance availability problems that had first plagued medical specialists in the mid-1970s re-emerged, and insurance for many other risks that had been covered was simply unavailable.

The "refusal" of insurers to write risks has been manifest not just in their withdrawal from some lines of insurance, but in the omission of

typically faced premium increases of 200% to 300%. Those day-care centers who were able to renew their policies faced premium increases generally in the range of 200% to 300%. The liability insurance premiums for aircraft manufacturers rose from $51 per aircraft in 1962 to $70,000 in 1985 with most of the increase resulting in the later years. Id.

18. Id. at 21. Note that the corresponding changes in claims are not included. There was an equally large increase in estimated losses from 1984 to 1985, id. at 20, which would appear to support the changing risks hypothesis, that the increase in premiums was justified by an increase in costs. But this is an illusion caused by an accounting bias, in which estimated losses are tied to premium income.

19. Id. at 20.

20. See infra notes 21 and 22 and accompanying text. Separate time series data on premiums levels, and quantities of coverage for property and casualty insurance are undefinable in any sensible way because of the heterogeneity of the insurance product.

21. TORT POLICY REPORT, supra note 1, at 10.

22. Id. at 6–15.
particular risks from insurance policies. Policies for day-care centers specifically excluded liability arising from child abuse by day-care employees; municipality insurance policies when available generally excluded liability risks from toxic waste dumps. In a pervasive trend, insurers are switching from conventional occurrence policies which cover liability for all damages occurring during the policy period, even if the damages are not apparent or claimed until years later, to claims-made policies which cover only liability from suits that are filed during the policy period. Claims-made policies thus exclude "tail coverage," coverage for risks incurred beyond the policy period.

Insurance for any risk is, of course, available for a sufficiently high price. In markets where the demand is very inelastic, because consumers are required by contractual obligations to carry insurance, one observes very high premiums instead of zero transactions. For example, firms specializing in the removal of asbestos from buildings paid insurance premiums in 1985 that were in some cases more than ninety percent of the coverage provided by the policies. The increase in premiums for insurance and the availability problems are aspects of the same phenomenon: a decrease in the supply relative to demand in the market. The decrease in transactions in some lines represents a decrease in the gains to trade in the market, rather than a refusal of insurers to write policies. It is this decrease in gains to trade that must be understood.

Another aspect is the adequacy of insurance and tighter limits on coverage. The limits on the amount of coverage that consumers could buy at the given premiums also dropped in 1985, in some cases to a fraction of the potential losses. The lower limits on coverage is not the same phenomenon as the decrease in availability. Insurers do not "ration" available capacity with the objective of an equitable distribution of capacity among their clients; they maximize profits. The tighter limits on coverage and the higher prices for excess coverage are examples of non-linear

23. For example, in 1985 the Edith B. Jackson child care program affiliated with Yale University obtained coverage that specifically excluded coverage for child abuse. The premium on this policy was six times the 1984 premium. Sorry, America, Your Insurance Has Been Cancelled, Time, Mar. 24, 1986, at 20.

24. Tail coverage could sometimes be purchased separately, but at a much higher premium. See, e.g., Blair & Makar, The Structure of Florida's Medical Malpractice Insurance Market: If It Ain't Broke Don't Fix It, 5 Yale J. on Reg. 427 (1988).

25. For example, Specialty Systems, Inc. of Richmond, Indiana paid $460,000 in insurance premiums for a $500,000 policy in 1985. Sorry America, Your Insurance Has Been Cancelled, supra note 23, at 18.

26. "Gains to trade" refer in this context to the benefits from exchange between risk-averse insurers and stock insurance companies, who are essentially risk-neutral because of their shareholders' opportunities to diversify in equity markets.
Insurance Market Dynamics

pricing of insurance coverage, not high supply prices, and represent a separate dimension of the insurance crisis.

A fourth aspect which must be accounted for is the higher industry profits. As illustrated in Table 2, over the period of the 1984 to 1986 insurance crisis, profit levels in the property-casualty insurance market improved greatly.\textsuperscript{27} Since the previous crisis in the middle-to-late 1970s, returns in the insurance industry had been declining—a period characterized in some of the industry trade literature as a "seven-year price war."\textsuperscript{28}

\begin{table}
\centering
\caption{Average Annual Rates of Return on Net Worth After Tax} \label{table:2}
\begin{tabular}{llll}
\hline
Year & Property-Casualty Insurance & All Industries & Difference \\
\hline
1976 & 10.0 & 13.3 & - 3.3 \\
1977 & 19.0 & 13.5 & + 5.5 \\
1978 & 18.1 & 14.4 & + 3.8 \\
1979 & 15.5 & 14.9 & - .4 \\
1980 & 13.1 & 14.4 & - 1.3 \\
1981 & 11.8 & 13.8 & - 2.0 \\
1982 & 8.8 & 10.9 & - 2.1 \\
1983 & 8.3 & 10.7 & - 2.4 \\
1984 & 1.8 & 13.5 & - 11.7 \\
1985 & 3.8 & 11.5 & - 7.7 \\
1986 & 11.6 & & \\
\hline
\end{tabular}
\end{table}

One of the most unusual aspects of the crisis was its catastrophic timing. The crisis hit the liability insurance market quite suddenly in late 1984 and early 1985 and followed a period of stable premiums. Any of the trends in market conditions to which the crisis is usually attributed—changes in tort law, introduction or revelation of new risks such as environmental risks and incidents of child abuse in day-care centers, or downward movement in interest rates—occurred over a number of years and not with the suddenness with which the crisis appeared. The increase in tort awards does appear to have accelerated recently,\textsuperscript{30} but not

\textsuperscript{27} Recall that commercial liability insurance, the lines of insurance subject to the crisis, constitutes only 10% of the property and casualty insurance industry.

\textsuperscript{28} "For the better part of seven years, the insurance industry has been engaged in a brutal price war." \textit{Tort Policy Report}, \textit{supra} note 1, at 22 (quoting Insurance Serv. Office, \textit{Financial Condition of the Insurance Industry—An Update} (1985)).


\textsuperscript{30} \textit{Tort Policy Update}, \textit{supra} note 1, charts I and J.
enough to account for the suddenness of the jump in premium levels and drop in availability from late 1984 to late 1985, following the period of relatively stable premiums since 1977.

The sixth aspect to be considered is the restriction of the crisis to particular lines of insurance. The recent insurance crisis, while more broadly based than the crisis of the mid-1970s, was restricted to commercial casualty lines. Automobile insurance, even automobile liability, did not experience the same dramatic changes.\(^{31}\)

Another feature of the liability insurance market is that in 1984–86 Canada experienced a crisis similar to the one in the United States. Yet many of the aspects of the American tort system that have been blamed for the insurance crisis and that are being changed as part of state tort reform are not features of the Canadian tort system.\(^{32}\) This raises the question whether the state reform of the American tort system is likely to have a significant impact on the performance of the insurance industry.

The eighth and final aspect is the recurrency of insurance crises, with the industry alternating between crises of high premiums/returns and periods of low returns. The crisis of 1984–86 was the most severe in the industry’s history, but it was by no means unique. From about 1977 the industry had experienced a period of stable premiums with declining returns, which was preceded by the crisis of the mid-1970s. That crisis was preceded in turn by a soft market that had begun in the late 1960s.\(^{33}\) The

---

\(^{31}\) The net premiums written for medical malpractice and other liability insurance jumped 80% from 1984 to 1985 (at the same time as availability was being cut back), whereas the net premiums written for total auto insurance rose 19%. The increase in premiums for automobile liability rose 21%. See Best’s Aggregates and Averages 1987, supra note 12, at 103.

\(^{32}\) See Angoff, supra note 5, at _____.

Some of the major differences between Canadian and U.S. tort law, as listed by Trebilcock, include the following: Canadian tort cases are typically determined by judges alone and not by juries; the contingent fee system for legal fees is not widely used in Canada and is prohibited in Ontario; awards for pain and suffering are subject to a judicially-imposed cap of less than $200,000 in Canada; punitive damages are rarely awarded in Canada; the movement to strict liability in the products context has not occurred in most Canadian provinces; employees in the U.S. can sue third parties, whereas in Canada they usually cannot; class action suits are subject to severe legal restraints in Canada. Trebilcock, supra note 9 (citing R. Pritchard, A Systemic Approach to Comparative Law: The Effect of Cost, Fee and Financing Rules on the Development of the Substantive Law, Working Paper No. 70, Civil Liability Program, Center for Studies in Law, Economics, and Public Policy, Yale Law School (1988)).

\(^{33}\) Barbara Stewart, writing before the recent insurance crisis, describes the profit cycle in the property-liability insurance market as alternating between “the competition phase” and “the crunch phase.” These phases correspond to the terms “tight markets” and “soft markets” used in the introduction to this Article. Her description of the crunch phase is worth noting:

The crunch phase of the profit cycle begins when profits are so squeezed that the continuing operation of most companies is threatened . . . . Typically, . . . the crunch is set off by a frightening external event or series of events, which by occurring when margins are already competitively beaten down, causes enough financial devastation to force a withdrawal from the market . . . . For example . . . in 1965 the scare was produced by the property damage and loss of reinsurance capacity following Hurricane Betsy. In 1974, a sudden erosion of surplus in a stock market collapse and a doubling of inflation were the principal causes of the crunch. The crunches of 1957 and 1969 in automobile insurance markets were set off by exceptionally
cyclical nature of the returns to insurers is puzzling when considered within the conventional economics of insurance markets.\(^4\)

II. The Insurance Crisis: Existing Theories

The debate about the insurance crisis of 1985-86 has been largely restricted to two hypotheses: recartelization and changing risks.\(^5\) The two hypotheses can be briefly characterized as follows: the cartel hypothesis says that the industry in 1985 recovered from a seven-year price war successfully re-establishing cartel prices;\(^6\) and the changing risks hypothesis says that because the liability insurance market is competitive, the insurance crisis must have resulted from a contemporaneous change in the insured risks, a change which was largely the result of an expansion of liability awards and standards in tort law.

Consider these hypotheses against the background of the empirical facts previously outlined. The cartel or conspiracy hypothesis is consistent with the increase in premiums, the increase in profits, and the suddenness of the premium increase. These implications explain the popularity of the conspiracy hypothesis in the press. It may also be consistent with the restriction of the crisis to some lines and the periodicity of crises. But it is inconsistent with the “adequacy and availability” features of the crisis. Firms do not, according to any known theory, establish a cartel to increase profits by completely withdrawing from a subset of segmented markets.

The cartel hypothesis is also inconsistent with the evidence on the structure of the insurance industry. Collusion is likely only in markets that have a small number of firms or are concentrated, with a cost disadvantage to smaller firms. In the United States property-casualty insurance market, there are more than 2000 firms, and, while most of these operate in only a small segment of the market, the costs of entry into other segments are minimal since the costs of entering the market, including estab-

---

\(^4\) Stewart, supra note 2, at 304-05. Stewart's theory of the profit cycle is based on market psychology, in contrast to the rational expectations approach in Part II, section C of this Article.

\(^5\) See infra Part III, section A.

\(^6\) The specific explanations of the crisis have been summarized exhaustively. For excellent examples, see Abraham, Making Sense of the Liability Insurance Crisis, 48 Ohio St. L.J. 399 (1987); Priest, supra note 9; Tort Policy Report, supra note 1; Tort Policy Update, supra note 1; Trebilcock, supra note 9.

\(^36\) A variant of this hypothesis is that the increase in premiums represented a collective effort to correct premiums that had been too low. See What Insurance Crisis?, supra note 6.
lishing a brand name and raising minimum capital requirements, have already been incurred by these firms.\textsuperscript{37} The changing risks hypothesis (that changes in premiums reflect changes in the basic product sold in the market) is at first blush very compelling. According to the conventional economic theory of insurance markets,\textsuperscript{38} a change in competitive premiums must reflect a change in the expected present value of risks.\textsuperscript{39} The changing risks hypothesis, however, is inconsistent with three of the empirical features of the crisis previously outlined. It cannot account for the suddenness of the premium increase in 1984 since the basic conditions facing the market, particularly the distribution of claims, developed over a number of years, not just in late 1984 and 1985. The implication that all changes in market premiums are due to contemporaneous changes in the rationally expected distribution of claims is implausible.\textsuperscript{40}

The changing risks hypothesis is also inconsistent with the increase in profits with the rise in premiums. An increase in the probability or expected size of claims should not result in an increase in profits. If the

\textsuperscript{37} The highest concentration in the property-casualty industry, as measured by the Herfindahl-Hirschman Index (HHI), was only 229 in 1985. \textit{Tort Policy Update}, supra note 1, app. at 9. The HHI is an index of market concentration that equals the sum of the squares of market shares, and varies between 0 (the limit of atomistic market structure) and 10,000 (a monopoly). Even for the Medical Malpractice line, which is relatively concentrated, the highest HHI was only 663. As the report states, these are low values for this concentration index and indicate that successful collusion among liability insurers is highly unlikely, even with the partial protection of the McCarran-Ferguson Act against anti-trust laws. \textit{Id.}

Abraham, supra note 35, at 402, expresses some (modest) sympathy with the conspiracy hypothesis:

Despite the weakness of the conspiracy hypothesis as a general explanation for the crisis, I think it likely that some companies took advantage of the crisis and raised premiums to levels higher than they would have raised them had there not been widespread turmoil. For example, a medical malpractice insurer which sells ninety percent of the coverage in a state with only two thousand physicians may be able to double its premium rates for certain classes of physicians without facing serious rate competition in the short run.

This argument confuses high prices, which can be explained by market power or monopoly, with increases in prices, which cannot. A conspiracy requires at least two large firms. In addition, the conjecture that firms raised prices more than if there had not been widespread turmoil (in particular, withdrawal of supply) is perfectly consistent with the assumption of competition in the liability insurance market.

\textsuperscript{38} \textit{See infra} Part III, section A.

\textsuperscript{39} The changing risks hypothesis is implied by the conventional economic model of a competitive insurance market previously outlined. Even if one extends this economic model to a still static model, but one in which premiums may reflect variance in claims as well as systematic risk in the Tort Policy Working Group's explanation, one is left with the theoretical implication that the increase in premiums must reflect, and be contemporaneous with, the rise in expected claims and uncertainty. \textit{See Tort Policy Update}, supra note 1, at 16-22 & app.

\textsuperscript{40} Insurance consumer groups have criticized the changing risks explanation of the insurance cycle on exactly this basis: "Today, as in 1976, the insurance industry blames the legal system for the huge premiums. Their implausible claim is that judges and juries became generous in 1976, stingy for the next eight years and inexplicably generous again in 1985." \textit{Hunter, Taming the Latest Insurance Crisis, N. Y. Times}, Apr. 13, 1986, § 6 (Business), at 3, col. 1. At the time he wrote the article, the author was president of the National Insurance Consumer Organization.
premium increases are due to some extent to a contemporaneous increase in the variance of the distribution of claims, then, as the Tort Policy Working Group report states, profits should decline.\footnote{41} In the popular debate between the two hypotheses—are the increases in premiums justified by the torts crisis, or has the torts crisis been exaggerated by the industry to rationalize excessive premiums?—the conspiracy hypothesis scores a point with the increase in industry profits.

Finally, an increase in the probability or expected size of claims due to developments in tort law or any other exogenous factor cannot explain the withdrawal of insurers from some lines of insurance, such as municipality and medical malpractice insurance. An increase in the size of a risk, in a standard economic model of an insurance market, raises both the supply and the demand curves for coverage, and does not result in a drop in the number of transactions.\footnote{42}

One version of the changing risks hypothesis which is close to the theme of this Article is offered by Trebilcock.\footnote{43} Trebilcock argues that the trend in the U.S. tort system towards the pursuit of the compensation or social insurance objective of tort law and away from the deterrence objective has made it increasingly difficult for insurers to predict risks. Trebilcock explains the liability insurance crisis as a response to this increased uncertainty.\footnote{44} In attributing the crisis to the developments in tort law in recent years, however, he does not explain why the liability insurance market consequences of the increased instability of the tort system were manifested suddenly in late 1984.

In short, neither the conspiracy hypothesis nor the hypothesis of a shift in underlying risks can account for the insurance crisis. Neither is consistent with all of the evidence. This Article’s criticism of the changing risks explanation of the insurance crisis is not with its premise that risks changed between early 1984 and late 1984, it is that this change cannot account for the dramatic jump in premiums and drop in availability.\footnote{45}

\footnote{41} Tort Policy Update, supra note 1, at 17 & app. The Update tests the implication of declining profits by comparing the financial performance of the industry over the 1967-80 period with its performance over the 1981-85 period, and finds that profits declined. The analysis and test seems to be directed at an explanation of the increase in premiums between these two periods, whereas in reality the market premiums were quite stable in the 1981-84 period, before the crisis in 1985. The report does not attribute the crisis in 1985 to the losses in the early 1980s. \textit{Id.} at 11 & app.

\footnote{42} The gains to trade between the consumer and the insurance company increase, rather than decrease, with an increase in the consumer’s exposure to risk.

\footnote{43} Trebilcock, supra note 9.

\footnote{44} Trebilcock’s thesis is that “the changing complexion of the United States tort system . . . explains many of the recent problems in availability, affordability and adequacy of liability insurance. Changes in parameters of liability and quantum of damage have made it increasingly difficult for insurers to price various types of risks.” \textit{Id.} at 929.

\footnote{45} Another hypothesis on the insurance crisis should be mentioned. The drop in interest rates in the early 1980s and the consequent decrease in investment earnings by the industry is often cited as a
An alternative hypothesis capable of explaining withdrawal from insurance lines is developed by George Priest. Priest argues that the insurance market cannot price risks based on full information about insureds, and therefore the market is vulnerable to the “unravelling” effect of adverse selection when even small shocks disturb the market equilibrium. When risk classification is not perfect, so that in a single risk group there are heterogenous risks, an increase in the premium charged will lead to the better risks dropping out of the market. This leads to a further increase in the premium to match the increased average risk of the group, which leads to a worsening of the selection of risks, and so on. In the extreme case, the market for some risks will disappear. Priest shows that markets for third-party liability insurance are much more prone to adverse selection with the expansion of liability standards and attributes the insurance crisis to a shift in tort law from first-party towards third-party liability.

Part II argues that the adverse selection effect will magnify the effects of a cycle on premiums and availability in insurance. As Priest so clearly documents, adverse selection is a fundamental feature of modern liability insurance markets. But the hypothesis that adverse selection per se led to the crisis of 1985 implies that profit rates should have dropped and that the rate of entry of capital into the industry should also have decreased. Neither of these predictions is supported by the evidence. In 1984 and 1985 when industry losses reached their highest level, the entry of capital into the market also peaked, as shown in Table 3. If adverse selection effects were responsible for the cut-back in supply in 1984 and 1985, one would have seen a withdrawal of capital from the market. Profit levels increased over the 1984–86 period, as shown in Table 2, and the stock market value of property-casualty insurers in the U.S. increased by 59.4% in 1985.

Contributing factor in the crisis. A decline in interest rates will increase the present value of claims and therefore increase competitive premiums. But this effect cannot account for the magnitude or the timing of the premium increase. Furthermore, a drop in interest rates raises both the supply and demand curves for insurance against future losses and therefore cannot explain why insurers withdrew from some lines.

46. Priest, supra note 9.
47. Id. at 216–26; see also Akerlof, The Market for “Lemons”: Quality Uncertainty and the Market Mechanism, 84 Q.J. Econ. 488 (1970).
48. Priest, supra note 9.
49. The Standard and Poor’s index increased by 26.9% over this period; insurance stocks are relatively low beta stocks. This is consistent with the hypothesis that the exit of reinsurance capacity in the U.S. market made the remaining capacity more valuable. See infra Part III, section B. Data are as of the second week in December. See Insurance Stocks, Best’s Rev.: Property/Casualty Ins. Edition, Jan. 1986, at 8.
Table 3

Net Capital and Surplus Paid-In
Property-Casualty Insurance Industry
1980-1986

<table>
<thead>
<tr>
<th>Year</th>
<th>Capital and Surplus Paid (Billions)</th>
<th>Net of Dividends to Stockholders</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>-1.495</td>
<td></td>
</tr>
<tr>
<td>1981</td>
<td>-1.775</td>
<td></td>
</tr>
<tr>
<td>1982</td>
<td>-1.248</td>
<td></td>
</tr>
<tr>
<td>1983</td>
<td>-1.945</td>
<td></td>
</tr>
<tr>
<td>1984</td>
<td>.322</td>
<td></td>
</tr>
<tr>
<td>1985</td>
<td>5.561</td>
<td></td>
</tr>
<tr>
<td>1986</td>
<td>1.300</td>
<td></td>
</tr>
</tbody>
</table>

III. The Dynamics of a Competitive Insurance Market

This Part develops a model of competitive insurance markets to explain the pattern of fluctuations in premiums and capacity in uncertain lines, such as liability insurance. The section begins with an overview of the conventional economic theory of insurance markets.

A. Background: The Conventional Economic Theory of Competitive Insurance Markets

Economic analyses of competitive insurance markets, in the absence of any imperfections such as asymmetric information, have been based on one of two models. The first, which I call the “classical” model of competitive insurance markets, is based on the assumption that the risks being insured are large in number and statistically independent. In this case, the law of large numbers essentially guarantees that a premium set slightly above the expected claim plus associated expenses will cover costs. Under the independence assumption, there are no problems of predictability of claims. Insurers are risk-neutral because of their abilities to pool risks and insurance premiums are simply reflections of the expected claims and associated costs. Any change in premiums, however sudden,

51. Independence means (more or less) that there are no common underlying factors linking the events of losses. Independence is a stronger condition than is necessary for the classical model of insurance. It is sufficient that the conditions of the weak law of large numbers hold, i.e. that the risks can be sequentially ordered in such a way that the dependence between risks approaches zero (at the same rate as 1/n) as the risks become far apart in the sequence.
must reflect a contemporaneous change in the present value of expected claims. A competitive insurance market will operate efficiently even in an economy without other well-developed capital markets, assuming there are no problems of asymmetric or hidden information.

Most explanations of the insurance crisis, including that developed below in Part III, sections B and C, attribute the crisis at least in part to a failure of the independence of risks in the market and the resulting difficulty insurers face in pricing and predicting risks. More specifically, blame is attached to the marked increase in the uncertainty facing insurers today as compared to historical periods with more predictable tort law and less technological uncertainty. But the independence of risks, while at the heart of the classical model of competitive insurance markets, would not be required for the efficiency of insurance markets if the insurance markets operate in an economy with well-developed capital markets.

The second economic model of competitive insurance markets, the “perfect capital market model,” does assume an economy with well-developed capital markets. In this model, an insurance corporation is simply mediating between consumers buying coverage, and the firm’s stockholders, who trade equity in a perfect market subject to no transactions costs, taxes, or informational asymmetries. Stockholders trading in such a market are averse to risks only to the extent that the risks are systematic, that is diversifiable in the capital market. Only risks that are correlated with the return on the aggregate stock market portfolio or with major economic factors such as the macroeconomic business cycle are systematic. Diversification within the insurance market alone is not necessary for insurers to be risk-neutral. The uncertainties faced by insurers are unlikely to be systematic because they are not linked to major economic factors. If the uncertainty facing insurers is the source of the insurance crisis, then the task is to explain why. As well as failing to explain the impact of uncertainty on the insurance market, the perfect capital market model is inconsistent with the insurance cycle in premiums relative to claims. In this model, premiums represent the expected present value of claims and associated expenses, just as the price of any security represents the present

52. Recent articles adopting the perfect capital framework are Fairley, Investment Income and Profit Margins in Property-Liability Insurance: Theory and Empirical Results, 10 Bell. J. Econ. 192 (1979), and Hill, Profit Regulation in Property-Liability Insurance, 10 Bell. J. Econ. 172 (1979).

53. For an explanation of the distinction between systematic and nonsystematic risks in the context of the perfect capital market model of an insurance market, see Hill, supra note 52; Fairley, supra note 52; Kraus & Ross, The Determinants of Fair Profits for the Property-Liability Insurance Firm, 37 J. Fin. 1015 (1982).

54. Furthermore, even when risks are systematic the additional premium to compensate stockholders is small if policy coverage extends only a few years. See infra Part III, section B.
value of claims promised by the security. Rational expectations rule out the possibility of cycles in any security markets, and insurance markets are not distinguished from other security markets in this model. The next section addresses the question of what truly distinguishes insurance markets from other security markets, in making them prone to cycles in earnings, prices, and availability.  

B. Capacity, Adjustment Costs, and Shocks in Insurance Markets

In conventional product markets, cycles, by which I mean fluctuations in prices relative to variable costs, are due to shocks in market capacity (or demand) combined with costs of adjusting capacity. My thesis is that the empirical dimensions of cycles in insurance markets are explained by the same logic as cycles in any other market. Developing this argument requires discussion of the relevant concepts of capacity, adjustment costs, and shocks to capacity.

1. Capacity

The characteristic of the liability insurance market that is most critical to the existence of cycles is that the average costs to insurers are highly uncertain, that is, the insured risks in the market are highly dependent. Insurers face four sources of uncertainty in the current economic environment: uncertainty in tort law, both in the determination of liability and the damages awarded, law with respect to the liability of insurers, uncertainty from natural or technological sources, and uncertainty about inflation.

Trebilcock offers an excellent synthesis of the developments in U.S. tort law that have added to the uncertainty insurers face in attempting to predict risks. One of the most prominent examples of these developments is the expansion of the scope of strict liability in the context of products liability. The movement towards strict liability in products adds

55. Another important aspect of actual insurance markets that the perfect capital market model does not explain is the existence of mutual insurance companies. If risks were independent, i.e. there were no underlying common economic factors in the distribution of claims across policyholders, as in the classical model of competitive insurance markets, then mutual and stock insurance companies would perform equally well in terms of pooling risks. But where there is any uncertainty at all in the frequency or size of claims, then mutual companies would not survive since they cannot insure against the uncertainty of common factors. Only stock insurance companies would survive. But mutuals do survive and so the model fails.

56. Trebilcock, supra note 9, at 929.

57. But cf. id. at 948 (citing Schwartz, The Vitality of Negligence and The Ethics of Strict Liability, 15 Ga. L. Rev. 963 (1981), as arguing persuasively that “the so-called ‘explosion’ of tort liability in the United States in the past quarter-century has been an expansion of negligence liability, rather than a substantial extension of the spheres of application of strict liability”).

469
uncertainty because it is unclear how far the expansion will continue in the future, because the principle itself is ambiguous, and, as Trebilcock states, "by imposing ex post liability for what may have been ex ante unknowable risks . . . courts have imposed retrospective liability for risks that neither manufacturers nor insurers were likely to have been able to predict at the time the product was introduced." 5

Insurers also face uncertainty in insurance law with regard to the liability of insurers. Uncertainty from natural or technological sources is exemplified by the future side-effects of drugs, future damages arising from the effects of municipal waste-dumps, and even predicting the frequency of child-abuse in day-care centers. A final source of uncertainty concerns inflation. The liabilities incurred by an insurer entering a policy are a mixture of real costs and, because of nominal limits to coverage, nominal costs. If insurers knew exactly the timing of future claims and could hedge perfectly against inflation, they could immunize their balance sheets against inflation and investment risk. Obviously, neither of these conditions is satisfied, and this type of uncertainty is significant.

These uncertainties, emphasized by almost every writer on the current crisis, are not in themselves sufficient to explain increasing premiums or insurers' reluctance to write some lines of insurance. None of these risks, except inflation, is systematic, that is, highly correlated with the return on the aggregate stock market portfolio, so the conventional economic theory previously outlined predicts that neither insurance company stockholders nor insurance managers should be averse towards them.

The important implication of uncertainty in explaining the cyclicality of the liability insurance market is that the net worth or surplus of any insurer with limited liability is a factor of production. Because companies do not know exactly what their costs are going to be, they must maintain a buffer of net wealth against adverse outcomes in order to avoid insolvency and in order to make their contractual promises credible. An insurer is constrained in the amount of insurance it can write at "reasonable" premiums by its quantity of surplus. Thus the relevant concept of

58. See Trebilcock, supra note 9, at 955.
59. Id. at 952 (emphasis in original).
60. For example, in the “triple-trigger” decision in Keene Corp. v. Insurance Co. of N. Am, 667 F.2d 1034 (D.C. Cir. 1981), cert. denied, 455 U.S. 1007 (1982), damages resulting from a tort in the distant past were found to be owed by the insurer at the time of the tort, the current insurer, and insurers in the period since the tort. In another context, Roberta Romano offers a lucid discussion of the uncertainty created by courts in essentially rewriting Directors and Officers liability insurance contracts. R. Romano, What Went Wrong with Directors and Officers Liability Insurance? (Jan. 1988) (unpublished manuscript on file with author).
61. See especially the excellent discussion in Trebilcock, supra note 9.
62. The additional need for surplus to meet regulatory standards is discussed in Part III, section E.
Insurance Market Dynamics

capacity in this market is the net worth of the insurer (alternatively labelled policyholders’ surplus or stockholders’ equity). The more uncertain is any line of insurance, the greater the net worth needed to write a given amount of insurance. In the perfect capital market economy that underlies the conventional theory of competitive insurance markets, the capacity constraint would be irrelevant. Because investors in the stock market of this ideal economy would be indifferent between holding securities directly in their personal portfolios and indirectly through insurance corporations, insurance corporations could increase their net worth at effectively zero net opportunity cost. That is, the cost of equity capital raised by a corporation to invest in securities would equal the rate of return on the securities. Net worth would still represent “capacity,” but the capacity constraint would never be binding because capacity would be costless. Capacity would expand beyond any constraint.

However, equity is not costless in reality, otherwise we would observe firms entering the industry with extremely large amounts of initial capital and maintaining enough equity relative to liabilities that bankruptcies would never be observed. As long as the probability of bankruptcy is positive, there is a benefit to issuing additional equity since any additional equity shifts risk-bearing from policyholders to the less risk-averse equity holders. A firm issuing additional equity would capture this benefit in the form of increased profits. That the benefit of increased equity is not fully captured is evidence of an offsetting opportunity cost.

The finance literature offers a number of reasons why equity financing is costly to a corporation, of which I shall consider the three most important. The disadvantage to equity financing most often discussed is the asymmetry in the tax code, whereby interest payments are deductible and

63. In applying this model, “capacity” is interpreted as the total capacity of all direct insurers plus capacity allocated to the North American market by the world reinsurance market. Reinsurance is the market in which insurers place some of their liabilities; to the extent that a direct insurer uses reinsurance, it is acting as a pure intermediary, passing on the policy liabilities to a final insurer. The decrease in capacity in the reinsurance market was even more severe than the decrease in capacity in the direct insurance market resulting from large losses and the bankruptcy of a number of large reinsurers and the withdrawal of reinsurers from the North American markets. See TORT POLICY REPORT, supra note 1, at 8.

64. Owners of a firm could increase the net worth of the firm by investing in securities through the corporation and financing that investment by shorting the same securities in their personal portfolios, leaving their contingent wealth positions unchanged.

65. Even if state guaranty funds bear the liabilities in the event of bankruptcy, the same argument holds. Regulators could insist on a surplus-to-premiums ratio of 100 to one instead of the current three or four to one, with no additional social cost.

66. See, e.g., R. BREALEY & S. MYERS, PRINCIPLES OF CORPORATE FINANCE 284-296 (2d ed. 1984); Myers, The Capital Structure Puzzle, 39 J. FIN. 575 (1984); Auerbach, Taxation, Corporate Financial Policy and the Cost of Capital, 21 J. ECON. LIT. 905 (1983). The costs summarized here are costs to initial shareholders of issuing equity and then investing the equity in financial assets.

471
Corporate earnings that are paid out as dividends are taxed once at the corporate level and again at the personal level, whereas earnings reaching security markets in the form of interest payments are subject only to personal taxes.

A second source of costs to equity financing is related to the asymmetry of information between managers of corporations (acting in the interests of current stockholders) and the capital markets, with respect to the future prospects of the firms. Managers of a corporation, through their knowledge of a firm and private knowledge of their own decisions, can estimate better than outside investors the future contingent returns from the corporation. Under this informational asymmetry, the willingness of managers to go to the equity market for capital, that is, the willingness of current stockholders to share the residual ownership of the firm with outside investors, is a signal that the future profits are relatively poor. The cost to a poor firm of sharing future profits with outsiders is less than the cost to a good firm, and in a rational expectations equilibrium the issuance of high amounts of equity is associated with poor firms. This makes the cost of equity high to any firm, since issuing additional equity reduces the relatively uninformed market's expectation of the firm's profits and hence reduces the price that it establishes for the firm's equity.

The third cost of equity lies in the agency problem inherent in the separation of ownership of the corporate assets by shareholders from the control of these assets by managers. Managerial decisions, in particular decisions about investment, generally will not be made in the best interests of every shareholder in a corporation. Agency problems represent a disadvantage to shareholders' investing in assets indirectly through a corporation rather than directly, and therefore constitute a cost to increasing the net worth of a corporation.

2. Adjustment Costs

In considering the costs of equity, it is useful to distinguish between the costs of maintaining any given level of equity and the costs of adjusting equity, that is, between costs that are incurred as a flow in proportion to

---

67. E.g., R. Brealey & S. Myers, supra note 66, at 400.
68. For development of this and related explanations of equity costs, see Leland & Pyle, Informational Asymmetries, Financial Structure and Financial Intermediation, 32 J. Fin. 371 (1977); Miller & Rock, Dividend Policy under Asymmetric Information, 40 J. Fin. 1031 (1985); Myers & Majluf, Corporate Financing and Investment Decisions When Firms Have Information That Investors Do Not Have, 13 J. Fin. Econ. 187 (1984).
69. See, e.g., Jensen & Meckling, Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure, 3 J. Fin. Econ. 305 (1976). Managers bear only part of the costs and benefits of investment and effort decisions, and therefore have incentives to take actions which are not in the collective interest of all shareholders in a corporation.
the stock of equity maintained by a corporation, and the costs incurred when the equity is increased by issuing more equity or decreased by issuing dividends or repurchasing shares.\textsuperscript{70}

The tax treatment of dividends, and the informational asymmetry between inside managers and the outside capital market, both lead to adjustment costs. As an illustration of the tax-induced adjustment costs, consider the following example. Suppose a stock insurance company decides to increase its net worth temporarily by raising one million dollars in the equity market today, and then lowering net worth by issuing one million dollars in dividends tomorrow. The costs incurred as a result of this endeavor would be the various transactions costs of issuing equity and dividends, and most importantly, the personal income tax liability on the dividends paid out. Even if there were zero transactions costs incurred by the firm, and therefore no adjustment cost to equity in entering the market, the tax on dividends would mean that capacity could not be lowered without incurring a tax liability.\textsuperscript{71} The market value of initial shares in the firm would fall by the amount of the tax liability as soon as the plan of temporarily raising net worth was announced.\textsuperscript{72}

Just as the tax liability leads to an adjustment cost of reducing net worth, the costs resulting from the information asymmetry discussed above are costs of increasing net worth. Only when a firm goes to the outside market to increase its net worth by issuing equity does it potentially incur the costs of selling its equity at an “under-valued” price. These costs would not be recovered if net worth is raised only temporarily. Thus, both

\textsuperscript{70} The distinction between flow costs and adjustment costs is critical for my analysis of cycles, as it is analogous to the distinction between the costs of maintaining and adjusting the stock of physical capital in the discussion of cycles in the production of goods. This distinction is not generally incorporated in the finance literature’s discussion of optimal capital structure because that literature is based on static models. Note that in the dynamic model discussed in Part III, section C positive adjustment costs are equivalently expressed as a cost advantage of retained earnings over external equity.

\textsuperscript{71} Note that this explanation is based on the asymmetry that dividends are taxed whereas purchases of equity in the market are not deductible. The asymmetry in the tax treatment of dividends and interest payments, i.e., the traditional double-taxation of corporate income, is not the basis of this adjustment cost argument. Under a realistic assumption that corporations must maintain a minimum payout ratio, the double taxation of corporate income implies a cost of maintaining a stock of equity.

\textsuperscript{72} This “trapped equity” effect of dividend taxation is well-known to financial economists. See Auerbach, supra note 66, at 926; Poterba \& Summers, The Economic Effects of Dividend Taxation, in RECENT ADVANCES IN CORPORATE FINANCE 227 (E. Altman \& M. Subrahmanyam eds. 1984).

This immediately raises two questions. First, is there really a personal tax disadvantage to dividends? I take the answer as yes, from the observation that taxes are paid on dividend income and the (admittedly somewhat weak) consensus that the penalty on equity returns in the form of dividends is positive. R. Brealey \& S. Myers, supra note 66, at 349, provide an excellent summary of the evidence on this penalty.

Second, if there is a tax penalty to the issuance of dividends, why do firms not avoid the tax liability by repurchasing shares instead? Firms could not redistribute cash to shareholders exclusively through share repurchase, since share repurchases are interpreted by the tax code as dividends if they are pro rata or regular. I.R.C. § 302 (1982). For a recent discussion of this issue, see Miller, Behavior Rationality in Finance: The Case of Dividends, 59 J. Bus. S451 (1986).
the tax liability and the informational asymmetry lead to adjustment costs; the informational asymmetry is a barrier to the entry of capital in the market, and the dividend tax is a barrier to exit.\textsuperscript{73}

These adjustment costs, however, represent costs of moving equity in and out of the entire corporate sector—not costs of mobility of financial capital across markets. That is, one must address the possibility of conglomerate corporations allocating equity to the insurance market through an insurance subsidiary during tight markets, and withdrawing this equity from the market when the market softens. While some of the capital inflow into the insurance market in 1985 and 1986 was in the form of cash injections from parent companies, this source of capital was not sufficient to eliminate the effects of the capacity constraint. There are two features of capacity and market organization that contribute to the mobility costs of capital within conglomerates. First, capital invested in other markets is transformed into physical capital, which is not fungible.\textsuperscript{74} Second, as a source of financial capital to the insurance industry, retained earnings from conglomerate parent companies is limited by the costs of establishing conglomerates, that is, by diseconomies to firm size. The limits on conglomerate size prevent equalization of the marginal return on retained earnings across all markets in the economy. If there were no such limits, a cycle of profits in the insurance industry could only mimic the rate of return to capital in the entire economy.

3. \textit{Shocks to Capacity}

The final component of the explanation of cycles is the nature of shocks to capacity. In the present context, the shocks are easily identified. The dependence or unpredictability of risks, which was the basis for identifying capacity as net worth,\textsuperscript{75} also provides the source of shocks. Any realization of the random frequency or size of claims that deviates from the statistical expectation is a shock to capacity. A high realization of claims, such as the market experienced during the early 1980s,\textsuperscript{76} will deplete capacity and a favorable realization will add to capacity.\textsuperscript{77}

\textsuperscript{73} The third cost of equity capital, increased agency costs, is a flow cost to maintaining a stock of equity, rather than an adjustment cost.
\textsuperscript{74} If capital were left in liquid assets there would be no advantage to withdrawing it temporarily from an insurance subsidiary.
\textsuperscript{75} See supra note 63 and accompanying text.
\textsuperscript{76} The high realization of claims is reflected in the low rates of return in Table 2. For specific examples of shocks to capacity, see Stewart, supra note 2.
\textsuperscript{77} An alternative source of shocks to capacity relative to demand in the market is, of course, fluctuation in the demand for insurance. The demand for insurance appears to be relatively stable, and an increase in demand is easy to rule out as the cause of the recent crisis.
Insurance Market Dynamics

A major shock to the North American insurance market in 1984 was the withdrawal of reinsurers from the market. Entry into the reinsurance market had been substantial during the late 1970s. In addition, manufacturers and groups of professionals such as doctors had formed captives, which are upstream insurance corporations established for the purpose of self-insurance. Under U.S. tax laws, captives benefitted from their status as insurance companies. By 1980, there were approximately 750 captives with premium income of sixty billion dollars.

The tax status of captives as insurers fluctuated during the late 1970s and early 1980s. To maintain this status, captives were forced to take on outside business, that is, to sell insurance to third parties, a condition they fulfilled by entering the reinsurance market. The tax benefits accruing to the parent company meant that the marginal cost of providing this reinsurance by captives was very low—resulting in deflated reinsurance premiums in the market of the early 1980s. In addition, reinsurers are not subject to the regulatory control that governs direct insurers and therefore are free to write larger amounts of business relative to net worth. The upshot was that the high claims in the early 1980s, combined with artificially low premiums, led to insolvency of many of the captives. Between 1984 and mid-1987, some ninety reinsurance companies left the market.

This section of the Article has outlined the concepts of capacity, adjustment costs, and shocks for competitive insurance markets under uncertainty in average claims. The implication of the adjustment costs and shocks, which is developed in the next section, is that the insurance markets will enter periods of high capacity and low premiums or periods of tight capacity and high profits and allow these periods to persist for some time. For example, suppose that the dividend tax is the only adjustment cost and that there are no barriers to the entry of capital. If the market enters a period of tight capacity because of unexpectedly large claims, it will issue some equity in response to the high profit rates in the market—but not enough to eliminate the capacity constraint. The capital market knows that once capital enters the market through the issuance of equity, the capital is trapped by the tax liability faced when it is removed.

78. Tax liabilities for insurers are lessened by conservative valuation of liabilities, and by the accounting practice of simply adding cash flows in a tax year, instead of discounting. When a business firm uses a funded retention program other than a captive to cover loss exposures, contributions to the retention fund are not tax deductible. Only losses incurred, such as liabilities to third parties, are tax deductible. Premiums paid to a captive by its parent, however, do qualify as a business expense (with limitations). For a review of the tax status and use of insurance captives, see Glen, Cooper & Porat, The Use of Captives in Risk Management, in 1 ISSUES IN INSURANCE, supra note 2, at 335.
80. Glenn, Cooper & Porat, supra note 78, at 382.
81. Insurance, supra note 79, at supp. 15.
82. Id. at supp. 11.
from the market through dividends. The capital market will be reluctant to allocate more equity to the insurance market if it expects periods of tight capacity and high profits to be temporary. Since high profits will add to the capacity through an increase in retained earnings, the expectation that the crisis is temporary will likely be realized. Situations of tight capacity persist in spite of the apparent free entry of capital into the market because the barrier to exit has the same effect as a barrier to entry.

C. The Basic Model

The adjustment cost argument can be illustrated with a simple economic model. Consider a competitive insurance market operating over time, in which during each period, for example, one year, a large number of consumers face risks of losses, which are indemnified by insurers. For simplicity, assume that the risks are the same for all consumers and that the size of the loss is known. Assume also, however, that the frequency of losses among consumers is highly uncertain even if the number of consumers is large. For example, consumers face a risk of liability for particular accidents and there is uncertainty in the standards of liability to be imposed by courts, but no uncertainty in the damages to be awarded. Finally, assume that the uncertainty in the frequency of losses is independent of aggregate risk in the economy, that is, independent of the market portfolio, so that stock insurance corporations are risk-neutral in the sense that they maximize expected profit.\(^8\)

The assumption of identical risks means that we can define the market for insurance, with one unit of insurance being defined as the payment of one dollar contingent upon the event of a loss to any individual. The short-run equilibrium of this market is described by the intersection of a demand curve for insurance and a short-run supply curve. The short-run supply curve for insurance is defined as the amount of insurance that the competitive market would supply at each premium, given the current stock of equity or net worth.\(^8\)

The demand curve for insurance is downward sloping, and depends on the expected frequency of losses. The supply of insurance in the short run depends upon the current stock of equity in the market, under the assumption that insurers have limited liability and face adjustment costs in changing this stock. Insurers are assumed to be averse to the possibility of

---

83. More precisely, risk-neutral stock insurers maximize expected market value in this model.
84. In describing the market as competitive, I am ignoring the heterogeneity of insurance firms. Capacity for the market is therefore the sum of capacities for firms, and the short-run supply curve is the horizontal sum of individual firm supply curves.
Insurance Market Dynamics

bankruptcy because it destroys valuable firm-specific assets such as the reputation of the firm.85

Note the shape of the short-run supply curve as shown in Figure 1. At low quantities, it will be almost perfectly elastic. Since the probability of bankruptcy is essentially zero, competitive insurers will supply coverage at an actuarially fair value, that is, at a premium equal to the expected present value of claims. The cost of insuring, up to the capacity constraint, is therefore proportional to the amount insured. At some quantity of insurance, the supply will become more inelastic as the market bumps up against its capacity constraint in the sense that any additional insurance makes the probability of bankruptcy significant. At this point the marginal cost of selling another unit of insurance is very high.

Figure 1

Short-run Supply Curve for Insurance Coverage

The capacity constraint in this market is peculiar, in comparison to conventional capacity constraints, in that it depends upon the market price.86 In particular, at a premium sufficient to cover costs even under

85. Alternatively, assume that state regulators prohibit any firm from selling so much insurance at any premium that the risk of its bankruptcy is more than some approved level. This means that the supply of insurance at a given premium will be constrained by the equity or capacity of the market.

86. In conventional product markets, a capacity constraint is determined by amounts of physical
the highest possible realization of the frequency of losses, the capacity constraint is never binding. At this high premium, revenues are certain to cover costs, and a stock of equity is unnecessary to avoid bankruptcy. This implies that the short-run supply curve is bounded from above. With very little uncertainty in the market, the supply curve will be nearly flat, whereas with much uncertainty, the maximum premium, \( P_H \), will be many times greater than the premium, \( P_L \), that is sufficient to elicit a small supply quantity. Holding constant the distribution of claims, an increase in the capacity shifts the supply curve to the right.

Figure 2 illustrates the short-run competitive equilibrium under the condition that capacity is at a level that yields a normal rate of return to capital in the market. The revenues in this case are sufficient to cover all costs, which include the claims and associated expenses, the expected bankruptcy costs, the fixed cost of maintaining the stock of equity, and a normal rate of return to capital in the market. If there were zero adjustment costs in the market, equity would be continually adjusted, and maintained at the level that corresponds to \( P^* \) in Figure 2. Figure 3 illustrates the short-run equilibrium under excess capacity and Figure 4 illustrates the case of a "crisis" of tight capacity and high premiums.

Figure 2:

Short-run Equilibrium with "Normal" Capacity

capital and does not depend upon the price charged in the market.
The dynamics of the model are described by a transition equation or relationship, which describes how the capacity (the "state variable") changes randomly over time. Capacity increases by the retained earnings plus equity issued in each period, which equals the realized profits minus dividends plus equity issued. When the market is in a "crisis" of tight capacity and high premiums, as in Figure 4, the equilibrium premium is high and the expected rate of profits and retained earnings are high. New
capital is rapidly entering the market, attracted by the high rate of profit. The equilibrium will move—stochastically—towards the equilibrium of Figure 2. When the market is in a state of excess capacity, as in Figure 3, economic profits are negative and the market may tend towards the equilibrium of Figure 4.

D. The Implications of the Basic Model

Some of the empirical, dynamic properties of insurance markets, including the facts outlined in Part I, section A of this Article, are explained by this simple model, and some can be explained by extensions of the model. I summarize the implications of the basic model below.87

Without adjustment costs, the market premium would always equal the expected present value of claims expenses, as in the conventional economic theory of competitive insurance markets. The basic prediction of the model is that premiums do not equal the present value of claims expenses; instead, the market fluctuates between soft markets of excess capacity with low premiums relative to costs and tight markets with high premiums.

The second prediction of the model is that "crises" of rapidly rising premiums tend to follow periods of stable premiums and low returns and result from cumulative losses in the market rather than contemporaneous shocks to the market. This prediction follows from the shape of the short-run supply curve (elastic at low quantities, inelastic at moderate quantities). With the market in a position of excess capacity as in Figure 3, a series of losses will not initially affect the premium. When capacity is sufficiently depleted, the inelastic part of the supply curve intersects demand.88 The model thus explains why crises are sudden relative to external shocks to the industry costs.89

One apparent implication of this explanation of crises is that the measured surplus in the market relative to demand should drop to a "critical" level before a crisis hits. This is usually measured empirically in the inverse ratio, as net premiums written relative to surplus. Figure 5 depicts this ratio over the period 1965 to 1985; note that before the crises of 1976 and 1985 this ratio rose sharply, exceeding 1.9 in both cases.

87. Section E of this Part describes the dynamic properties of the market that are not explained in the simple model together with the necessary extensions of the model.
88. In this simple model, a cumulative loss follows from a sequence of bad draws by the market. But this is just an artifact of the simplifying assumption that losses are insured and realized in the same period. See the extension discussed in section E of this Part to multi-period, overlapping insurance contracts.
89. Further discussion of the suddenness or catastrophic character of crises is offered infra section E of this Part.
The hypothesis that crises are caused by capacity constraints, as in this explanation, has been criticized by many observers on the grounds that this ratio is well below the level of three or four considered safe by state commissioners monitoring the financial solvency of insurers. This criticism would be valid if the actual market matched the model’s assumption of firms selling a homogenous product in a perfectly competitive market. In

90. A.M. BEST CO., BEST’S AGGREGATES AND AVERAGES: PROPERTY CASUALTY (1965-1985). Revenue, or net premiums written, is endogenous and the interpretation of this ratio as surplus to exogenous demand—a correct measure of the relative capacity of the market—is not accurate. The ratio is at best a rough indication of relative capacity over periods of stable premiums. In particular, the jump in the ratio in 1977 or 1986 is not an indication of precipitous drops in relative capacity. 91. E.g., Priest, supra note 9.
that case, the aggregate capacity-to-demand ratio would convey the same information about capacity constraint that the ratio for a single firm would. But in the actual market, firms are heterogeneous, and in particular the ratio of surplus to revenue varies across firms. A ratio of two for the aggregate ratio implies that the marginal firms, those closest to a capacity constraint, have a ratio much larger. Furthermore, a large increase in the costs of the marginal or highest cost firms will drive up the equilibrium price in a competitive market, since the increase would drive up the industry marginal cost curve which is the short-run supply curve.

The implication is that a better test of the capacity constraint hypothesis is whether the marginal firms in the market experienced capacity problems. One indication of this is the number of insolvencies in the market just before and during a crisis. The number of firms on the “wrong side” of the capacity constraint is surely highly correlated with the number of firms near the capacity constraint. Table 4 lists the numbers of insolvencies from 1969 to 1987. These data offer strong support for the hypothesis of a capacity constraint, since the number of insolvencies at least quadrupled just prior to the crises in 1984 and 1976.

Table 4

<table>
<thead>
<tr>
<th>Year</th>
<th>Insolvencies</th>
<th>Year</th>
<th>Insolvencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>1969</td>
<td>1</td>
<td>1979</td>
<td>3</td>
</tr>
<tr>
<td>1970</td>
<td>4</td>
<td>1980</td>
<td>4</td>
</tr>
<tr>
<td>1971</td>
<td>8</td>
<td>1981</td>
<td>6</td>
</tr>
<tr>
<td>1972</td>
<td>2</td>
<td>1982</td>
<td>9</td>
</tr>
<tr>
<td>1973</td>
<td>2</td>
<td>1983</td>
<td>4</td>
</tr>
<tr>
<td>1974</td>
<td>5</td>
<td>1984</td>
<td>20</td>
</tr>
<tr>
<td>1975</td>
<td>20</td>
<td>1985</td>
<td>23</td>
</tr>
<tr>
<td>1976</td>
<td>6</td>
<td>1986</td>
<td>22</td>
</tr>
<tr>
<td>1977</td>
<td>6</td>
<td>1987 (To July 1)</td>
<td>10</td>
</tr>
<tr>
<td>1978</td>
<td>6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A series that is equally relevant is the number of firms on the “watch list” of the National Association of Insurance Commissioners. This list covers those insurers considered to be in potential financial difficulty. The percentage of insurers on this list was 23% in 1985 and 24% in 1986—double the 1983 figure of 11.5%.

The third prediction of the basic model is that if the possible variation in capacity (dedicated to a particular line of insurance) is severe and demand inelastic then the market will tend to be more often at high premiums or low premiums than near a normal rate of return. More precisely, the long-run price distribution will be double-peaked, above and below the premium that yields a normal rate of return to capital in the market.

This prediction follows from the logistic shape of the short-run supply curve. As capacity and the supply curves of Figures 2, 3, and 4 vary stochastically over a wide range, and if demand is fairly inelastic, the intersection of the demand and supply curve will most often be at a price near $P_H$ or near $P_L$, and not often near $P^*$. The prediction of a double-peaked distribution of prices over the long run is an explanation of the apparent “instability” of competitive insurance markets in the face of uncertainty—that is, the tendency of these markets to be either soft with low premiums and low returns or tight with high profits and premiums.94

To restate this prediction in terms of the market price dynamics, insurance markets with uncertain losses face the following dilemma. When premiums are very high or very low, the premiums tend to be stable. But near $P^{**}$—the premium needed to generate a normal rate of return in the market—the premium tends to bounce away from its current value; because the short-run supply is relatively inelastic near $P^*$, only a small change in capacity (a small deviation in profit from its expectation) is needed to induce a relatively large change in the equilibrium premium.

A fourth prediction of the basic model concerns the persistence of soft markets relative to tight markets. The capacity level at which the market converges in expectation is higher than the level that guarantees a normal rate of return to capital as depicted in Figure 2 (the level that would be maintained if there were no adjustment costs). If the market is currently at this “normal” capacity, economic profits are zero. But the stock of surplus, invested in securities, earns a normal rate of return and continues to grow at this normal rate of return (multiplied by one minus the dividend payout ratio). The expected growth in surplus tends to push the market towards the equilibrium of excess capacity as in Figure 3, and once the market is in the stage of excess capacity, it will tend to stay there. In contrast, in a tight market period the internal contribution to capacity is the sum of positive economic profits plus a normal rate of return on the invested capital and the movement out of a tight market is more rapid. In a tight market, even if the claims are realized at their expectation the market will thus drift towards excess capacity, whereas to take the market

94. The condition that capacity vary by extreme amounts is most realistic for some lines in a multi-line extension of the model. See infra section E of this Part.
out of a period of excess capacity a poor realization of claims relative to market expectations is necessary. 95

The fifth prediction is that greater uncertainty in claims leads to greater variance in the amplitude of insurance market cycles. An increase in the uncertainty of claims96 in this model has the effect of making the short-run supply curve more inelastic at any premium, which makes the premium more sensitive to a change in capacity. Thus an increase in uncertainty has a double-barreled impact on the variance of premiums over the long run in a competitive insurance market with costly capacity. Not only are the shocks to capacity larger but the impact of any given shock on premiums is magnified as well. The increase in the severity of the insurance cycle since the 1960s can be attributed to the increase in the uncertainty in the environment facing the insurance market. This increase in all four sources of uncertainty outlined in section B of this Part is the dominant feature of insurance market conditions in recent years.97

This Article has emphasized that a competitive insurance market facing a severe but constant amount of uncertainty will experience periodic crises, and therefore that a dramatic change in premiums over a short period may not correspond to a contemporaneous shock to the distribution of risks in the market. Consideration of the short run impact of a sudden increase in uncertainty is nonetheless a useful result of the model in understanding the industry. A sudden increase in uncertainty of claims in an insurance market that is in or near a capacity constraint will result in an increase in the margin of premiums over expected claims. The effect of this increase in a capacity-constrained market is to make the supply curve steeper. This drives up the equilibrium price.

This comparative static result could explain the insurance industry's strong reaction to even single tort decisions.98 The model predicts that if the market is in or near a capacity-constrained equilibrium, then premi-

95. The prediction that crises are only temporary is consistent with the current evidence that premiums have levelled off and availability improved in 1987.
96. As used in this Article, "an increase in uncertainty" is defined as a mean-preserving spread in the claims distribution.
97. The second implication of the comparative static result is the following cross-sectional prediction which explains why the crisis has been restricted to commercial casualty rather than automobile insurance and other personal lines. The lines with the most uncertainty experience the most severe cycles. This prediction comes closest to explaining real-world actions in the multiline insurance market in which capacity at any time is allocated across the lines. See infra note 101 and accompanying text.
98. In the Canadian industry, the rise in premiums in municipal liability insurance followed that in America. It seemed to be a reaction to McErlean v. Sarel et al., 32 C.C.L.T. 199 (1985). In McErlean, a boy was awarded damages of 6.3 million Canadian dollars, an amount huge by Canadian standards, for injuries sustained in a dirt bike accident while riding illegally on municipality property. The decision was later overturned, both on liability and quantum of damages, 42 D.L.R. (4th) 577 (Ont. C.A. 1987).
ums will tend to react to changes in the worst-case scenario, rather than changes in the expected claims. An increase in the upper tail of the claims distribution will raise premiums even if the change in expected claims is small.

An increase in uncertainty of common factors from 1983 to 1984 could account for all of the empirical features of the liability insurance crisis that are outlined in this Article, including the increase in premiums, the cut-back in insurance availability, the increase in profits and the entry of new capital into the industry, except for the suddenness of the crisis relative to changes in tort law. Undoubtedly, an increase in uncertainty was a major contributor to the crisis, but the depletion of industry capacity through an accumulation of losses is necessary to explain suddenness of the crisis relative to changes in exogenous market conditions.

E. Extensions and Additional Implications

The basic model of a capacity-constrained insurance market outlined in section C of this Part is extended in this section to explain more fully the particular vulnerability of the insurance market to fluctuations in capacity, as well as several additional features of the insurance crisis. Can adjustment costs or rigidities in financial capital explain the extent of the crisis? On the one hand, the crisis in the supply of liability insurance in 1984-1986 was much more severe than the temporary supply shocks in other markets. On the other hand, capacity, which is financial capital, should be much more flexible than capacity in other industries, notwithstanding the adjustment costs delineated in section B of this Part. Financial capital is shifted relatively easily across insurance markets, whereas in manufacturing industries, for example, because capacity is physical capital, it is not easily adaptable to other uses and therefore less mobile.

99. A surprising corollary is that the increase in the premium-expected claim margin with a new worst-case scenario can benefit the competitive industry by increasing expected profits on new policies. An increase in uncertainty raises marginal costs, hence premiums, more than average costs. Of course, in most cases, this benefit will be more than offset by the increase in the expected costs of covering claims for past policies to which insurers are committed.

100. Stewart attributes the insurance cycle to this mobility of financial capital and argues that financial capital can respond quickly to changes in market psychology: "Because insurance [capacity] is both financial and psychological, changes in the supply of insurance can occur as quickly as money can be transferred and can be as volatile as a state of mind." Stewart, supra note 2, at 291. This is exactly the opposite of the explanation offered in this Article. If capacity were subject to zero adjustment costs, then there could be no fluctuations in premiums relative to the expected present value of claims. Stewart's analysis leads her to state: "Restricting freedom of entry and exit from insurance markets is a possible way to reduce competition and thereby to mute or avoid profit cycles." Id. at 313 (emphasis supplied). Stewart concludes, however, that such restrictions are not in the public interest. Id. at 316.
1. The Suddenness and Severity of the Insurance Crisis: Additional Factors

Three additional structural and regulatory features of the insurance market make it sensitive to even small rigidities in capacity.

a. Multiple Lines

Consider an extension of the model to a market with many lines of insurance. Suppose that the lines differ in the degree of uncertainty in the frequency of losses. Suppose further that the frequencies of losses are perfectly correlated across lines allowing us to abstract from the incentives to pool risks across lines.

Under the perfect correlation assumption, a "separation theorem" allows us to consider the short-run equilibrium attained at a particular point in time. First, the total capacity available in the market is allocated to the various lines of insurance. Second, the equilibrium in each line, conditional upon the capacity allocated to it, is attained as presented in the model of the previous section in Figures 2, 3, and 4. Initially, capacity is allocated so as to achieve equality in the shadow price, or value per unit, of capacity across lines. The equilibrium is established, losses are realized, and the resulting total stock of equity grows at a normal rate of return to yield the initial stock of equity for the next period in the model.

In the short-run equilibrium of this model, capacity may be binding or nonbinding. If capacity is binding, the allocation will result in the highest premiums for the most uncertain lines. The most uncertain lines bear the brunt of a negative shock to aggregate capacity for the entire market. This effect is illustrated in Figure 6. With variations in the total amount of capacity, premiums in the lines with predictable risks will fluctuate only a little; for these lines equity is a less important input into the production of insurance, and variation in its shadow price over time will matter little. For less predictable lines, premiums fluctuations will be severe.101

101. When capacity is plentiful and the shadow price on capacity is small, more capacity will be dedicated to uncertain lines since these lines require more for the offer of any given amount of coverage. When capacity is scarce, uncertain lines will be allocated very little capacity.
Thus, when capacity is mobile across lines the most uncertain lines, such as liability insurance, are the "residual lines" into and out of which capacity flows most rapidly. These lines will experience the most severe cycles, both in magnitude and suddenness of the timing. The multiproduct nature of the insurance market is critical in explaining the magnitude of the effect of capacity cycles for these lines.

b. **Adverse Selection Effects**

Any cycle in premiums will be magnified by adverse selection. To this point in the discussion and the model, I have assumed that all risks in a given pool, charged the same premium, are the same. Operationally, premiums and the categorization of risks can depend only upon observable characteristics of insureds. However, potential insureds may have additional information that is relevant to predicting losses. The consequence of
this asymmetry in information\textsuperscript{102} is that insureds with different risks face the same premium, which reflects the average risk in the pool. The better risks in the pool—those individuals whose hidden characteristics are “good”—will demand less insurance and may even drop out of the market rather than pay what for them is an inflated premium while the poorer risks will find the premiums charged a bargain and buy more insurance. The average risk in the pool is increased because of adverse selection. This means the premium must also be raised, which in turn causes better risk insureds to drop out of the market. In the most extreme cases, the market will disappear.

In much the same way, an increase in the premium because of a negative shock to capacity will lead to an increase in the relative number of good risks dropping out of the market or decreasing their demand for insurance. The worsening selection of risks will cause the market premium to rise even further. The total increase in premium will be greater because of the adverse selection effect.

This effect is illustrated for a competitive insurance market in Figure 7.\textsuperscript{103} The short run supply price, facing a particular risk category, is denoted by a curve \( P(R) \) where \( R \) is the average risk of the category. The average risk and the market price are determined in equilibrium as in Figure 7. The risk group is assumed to be very small relative to the demand for the entire market, so that the supply curve is essentially perfectly elastic and the dependence of supply price on quantity is ignored. \( P(R) \) is upward sloping since a higher average risk of the group is met with a higher premium.

A second curve depicts the consumer's side of the market and the selection of risks in particular. Upward sloping \( R(P) \) is the average risk of policies demanded as a function of the premium charged the risk group. (For comparison, the selection of risks in the absence of adverse selection is also depicted. In this case, the average risk is fixed at \( R \).)

\textsuperscript{102} As used in this Article, “asymmetry in information” does not necessarily imply that individuals are better informed than insurers about risks. Instead, it implies that individuals have some information about probabilities and likely amounts of losses that insurers do not have. The converse may also be true, i.e., the information sets may be nonnested.

\textsuperscript{103} One of the effects of adverse selection in a competitive insurance market is that in equilibrium a nonlinear price schedule is charged for coverage. See Rothschild & Stiglitz, \textit{Equilibrium in Competitive Insurance Markets}, 90 Q. J. ECON. 629 (1976). Individuals with higher probabilities of losses self-select into contracts with higher coverage and premiums. I ignore this implication, imposing an ad hoc constraint that premium schedules be proportional to coverage, in order to illustrate clearly the destabilizing effect of adverse selection.
Figure 7
The Impact of Adverse Selection on the Insurance Cycle

The equilibrium in the market is at the intersection of the curves \( R(P) \) and \( P(R) \). A shock to the market shifts the supply price function \( P(R) \) upward. A given shock has a bigger impact on the equilibrium under adverse selection and is greater the more severe the adverse selection, that is, the greater the slope of \( R(P) \). If adverse selection problems are severe in the insurance market even modest shocks and adjustment costs can lead to large swings in premiums and availability.

Priest has argued that the shift in tort law towards third-party liability from a first-party liability system has caused the apparent instability of the insurance market by making risk classification more difficult, introducing adverse selection effects.\(^\text{104}\) The argument that the risk classifica-
tion problem in third party liability insurance is especially severe is very persuasive, because of the existence of asymmetric information in the market. For example, producers who are in the market for products liability insurance are more idiosyncratic than the consumers of their products. They are also likely to be much better informed about the riskiness of their products than are their insurers, since the cost of this information to the producer itself is lower than to its insurer, and, in addition, this information is of value to the firm independent of its interaction with insurers.

Figure 7 shows, in the context of the capacity cycle hypothesis, that the slope of $R(P)$ would be greater for third party liability insurance markets than for first party insurance. As a consequence, the cycle in premiums is of greater amplitude. The cycle in premiums in the insurance market has become much greater since the 1960s. The analysis here supports Priest's theory that much of this increased instability can be attributed to the greater adverse selection effects as the tort system moves towards expansion of third party liability.105

\[ \text{c. The Cyclical Impact of Solvency Regulation} \]

One of the purposes of the state regulatory systems in the U.S. insurance industry is to monitor and ensure the financial viability of insurance companies. Insurance commissioners monitor companies most closely in the tight market stage of the insurance cycle, when the rate of insolvency is highest.106 I consider here the possibility that solvency regulation in its current form may exacerbate the insurance cycle.

Insurers are monitored on the basis of a number of financial ratios. The most important is the ratio of the flow of net premiums written (revenue) to the stock of surplus. This ratio is restricted, implicitly in some states and explicitly in others, so that it does not exceed three or four. The constraint is revenue/surplus \( \leq 4 \).

105. Id. at 1566, 1585.

106. The public interest explanation of solvency regulation is found in monitoring incentives. If left to consumers, monitoring would be insufficient. Three problems lead to this market failure. First, investment by many consumers in the information about the solvency of firms means that information expenditure would be duplicated. Second, each consumer has an incentive to free-ride on the monitoring by other consumers; if other consumers are monitoring the quality of insurers, only safe insurers will survive in the market, reducing the incentive for any particular consumer to monitor. Third, and most important because it distinguishes the market failure in this context from the problems of market provision of product quality in general, is an agency problem. See Mayers & Smith, Contractual Provisions, Organizational Structure, and Conflict Control in Insurance Markets, 54 J. Bus. 407 (1981). Insurance companies have an incentive to undercapitalize, since previous purchasers of insurance from a company, who have an interest in its continued viability, are not represented in the capitalization decision. The private contracting solution to this problem would in theory be a requirement that an insurer maintain a deep and adequate margin of surplus. Solvency regulation may be the most efficient means of enforcing this requirement.
Insurance Market Dynamics

The regulation effects both the amplitude of the cycle and the suddenness of crises. Regarding the first effect, if the demand for insurance is inelastic, then revenue is pro-cyclical with the premium cycle in the sense that revenue is highest when premiums are highest. The demand for liability insurance is inelastic. Therefore, the shadow cost of the regulatory constraint is also pro-cyclical. Since the premium equals the marginal cost of insuring the regulation leads to greater fluctuations. During periods of tight capacity, the regulation induces both a higher premium and additional issuance of equity.

The second effect of solvency regulation is on the "catastrophic" nature of crises. When firms bump up against the capacity constraint during a period of low premiums and returns, they are induced to cut back on the quantity of insurance offered. This causes the market premium to rise. At the higher premium, when demand is inelastic the amount of insurance that the constrained firms can offer is reduced by regulatory constraint, and insurers are induced to cut back even further. This perverse effect can suddenly turn a modest cost increase into a large increase in premiums.

Figure 8 illustrates this point in the competitive market model. The nonregulated supply curve, $S$, is upward sloping in this market. The regulatory constraint is given by the condition that revenue divided by surplus not exceed 4. This is represented in the figure by the inequality \( \frac{PQ}{\text{Surplus}} \leq 4 \), where $P$ is the premium, and $Q$ the quantity of insurance transacted. This constraint is a downward sloping iso-revenue curve, which forms part of the short-run supply curve. Thus, the regulatory constraint, RBE, induces a backwards-bending supply curve: at any premium higher than $B$, firms will choose to sell at the market premium up to its constraint, and the supply curve under regulation is described by the curve ABR.

107. The rise in revenue in 1985 with increases in premiums and cutbacks on availability can only be explained by using demand inelasticity. However, the literature does not contain any econometric estimates of the demand for liability insurance.

108. The additional equity is carried into the soft market periods resulting in a lower premium in these periods than in an unconstrained market.

109. In modelling the effect of this regulatory constraint, I am assuming for simplicity that the entire market is governed by a single regulation, rather than one for each state. I also ignore the reinsurance market. Insurers can circumvent the regulation to some extent by passing on the risks to reinsurers, who are not bound by the regulation. The regulation will have an effect providing that reinsurance is not provided in infinitely elastic supply, a condition which is obviously realistic.

110. Given the market premium, a competitive, price-taking firm will offer coverage up to the point where the regulatory constraint is binding, rather than offer more coverage at a lower price. The latter strategy would raise costs without raising revenues.
The implication of the backwards-bending supply curve is that when demand intersects the natural supply curve at a quantity near, but to the left of, B, a slight decrease in the capacity will induce a large jump in the equilibrium premium. The equilibrium premium will increase to \( P_2 \) from \( P_1 \). The current form of the solvency regulation may not only add to the magnitude of the cycle, but also to the catastrophic timing of crises.

The hypothesis that the cycle is caused by the regulatory constraint is unlikely. The hypothesis that regulation has contributed to the cycle is a more realistic possibility. One possible remedy to the perverse effects of this regulation would seem to be to emphasize the ratio of the stock of liabilities to surplus, rather than the ratio of the flow of revenue to surplus. But such an approach would be problematic given current reporting procedures because insurers have wide scope in defining liabilities. For example, insurers have been criticized for over-estimating the reserve liability that represents the present value of future claims from policies beyond the policy period—the "Insured But Not Reported" reserve. Statutory restrictions on liability accounting restrict the relationship between liabilities and revenue. For example, under the "loss-ratio" method of re-
serve calculation, the loss reserve must be no less than sixty percent of premium revenues net of losses incurred, for the first three years of a policy.111 This monitoring criterion ties to revenues and leads to the same effect as the solvency regulation that now exists.

2. The Withdrawal of Insurers from Some Lines

The multiple-product nature of the market also explains the withdrawal of insurers from some lines. In the multiple lines model above, when the shadow cost of capacity is sufficiently high, the supply price at zero quantity will exceed the maximum demand price for some lines. The lines with zero transactions will be the most uncertain lines: at a given level of capacity, the supply curve is higher for uncertain lines.

This explains the puzzle of why there were no gains to trade in some lines in 1985, while there were gains to trade in early 1984—even without a dramatic change in exogenous market conditions over this period. It is tempting to conclude that because crises are temporary—an implication of the capacity-constrained model—availability problems eventually will be resolved for all lines. An implication of the model, however, is that if the frequency of risks in a particular line is extremely uncertain, then there are no gains to trade even in soft markets. An increase in the uncertainty in some lines such as medical malpractice insurance may have led to an indefinite withdrawal, not a cyclical withdrawal, of the supply of stock insurance.112

3. The Tighter Ceilings On Coverage

In the simple model, all contracts are uniform pricing contracts; a consumer can buy as much coverage as desired at a single price. The simple model does not attempt any explanation of ceilings on coverage and the sharp drop in the maximum coverage that these contracts have contained. The element missing from this simple model is uncertainty in the average size of a loss (conditional upon the event of a loss), as contrasted with uncertainty in only the frequency of losses. For example, tort law is a major source of uncertainty, but the simple model assumes uncertainty only in the standards of assigning liability, ignoring uncertainty in dam-

112. Some of the withdrawal of stock insurers will be met with the entry of mutual companies. To this point the role of mutual insurers has been ignored in this theory of insurance market dynamics. In markets where risks are dependent, mutual insurance is inherently a second-best form of insurance. A mutual contract protects against individual-specific risk but does not offer any insurance against common factors in the distribution of losses. While mutuals do not bear all of the costs of capacity that a stock company does, their inherent disadvantage in the allocation of risk-bearing restricts their market share to that left by the limited capacity of stock insurers.
ages awarded. As I have shown in a previous paper, with this second type of uncertainty and a constraint on capacity in an insurance market, the equilibrium insurance contracts cover all losses up to a particular limit. The explanation of limits on coverage is simple. The correlation among the size of losses across individuals means that the opportunity cost of offering one more dollar of coverage to an individual who has a large policy is greater than the cost of the first dollar of coverage; the shadow cost of the capacity constraint enters the marginal cost of coverage with greater weight at higher coverage. A ceiling on the coverage of each policy offered protects the insurer against bankruptcy in the event of very large average losses. This effect overwhelms the fact that a risk-averse individual most wants to insure against the largest losses.

IV. Policy Implications

The analysis in this Article has implications for four types of public policy and regulation for the insurance industry: solvency regulation, government policies to increase insurance capacity, premium rate regulation, and the repeal of the insurance industry's exemption from federal antitrust statutes.

A. Solvency Regulation

This Article has suggested that solvency regulation as it is currently implemented can exacerbate any cyclicality of the market. A case can be made for regulatory monitoring and assurance of adequate surplus by firms in the market, but there is a need to develop new solvency parameters to replace the premiums to surplus ratio. Proper measures of solvency must reflect the increase in solvency that follows from an increase in rates.

B. Government Policies to Increase Insurance Capacity

Policies that have support in economic analysis include increasing availability of insurance through government support for insurance exchanges. Government support can be advocated on the basis of the public good derived from the exchanges. To the extent that the capacity

114. See supra text accompanying notes 106–11.
115. See supra note 106.
Insurance Market Dynamics

crunch in the industry is in part a tax distortion, action to increase capacity can be justified on second-best grounds.

A related policy that also has direct implications for increasing availability of insurance has been implemented by many states since the start of the 1985 crisis: the authorization or development of self-insurance programs for certain categories of insureds. The Tort Policy Working Group reports that most of the provisions deal with self-insurance programs or the formation of risk-pools for municipalities or other public entities, but with the 1986 amendments to the Products Liability Risk Retention Act of 1981, the option of self-insurance or insurance purchasing groups increased for private entities as well. The removal of regulatory impediments to self-insurance or other alternatives is clearly an appropriate policy response to the insurance crisis. This policy would mitigate the effects of tight capacity that are analyzed in this article.

C. Premium Rate Regulation

No economic justification can be found for the increased insurance rate regulation that the crisis has provoked in many states. Following the recommendation of the Cuomo Commission, New York established a "flex-rate" system in 1986 whereby prior approval by the insurance commission is required for any rate increase in excess of twenty-five percent. Flex-rate regulation represents a price ceiling, and possibly a price floor, in an insurance market. If the thesis of this paper is correct, then the effect of the flex-rate regulation, to the extent that it is binding, is that of a price ceiling in a competitive market. The effect of a price ceiling in a competitive market is well-known: it exacerbates the capacity problem in the market by discouraging the entry of new capacity.

In the context of insurance markets, two factors magnify the detrimental effect of a price ceiling on capacity. First, because a price ceiling reduces the flow of profits, it has a direct negative impact on future capacity, which here is the cumulation of retained earnings as well as new

117. See supra Part III.
118. TORT POLICY UPDATE, supra note 1, at 73.
120. Previously, New York law provided for a "file and use" rating system for commercial lines of insurance such as most liability lines. The Commission argued that an effective regulatory approach should "regulate both upward and downward fluctuations of . . . rates," responding to "unwarranted rate reductions in soft markets, which lead to price shock in hard markets such as the one being experienced today." CUOMO COMM'N, supra note 2, at 93-100 (emphasis supplied).
121. See TORT POLICY UPDATE, supra note 1, at 74.
122. If the regulation never prohibits premium increases, then the only effect of the regulation is the administrative cost, which is significant in itself.
capital. Second, as shown in Figure 1, the short run capacity constraint in this market depends on the premium level as well as the capacity level so that the immediate short run supply is not so inelastic as in conventional product markets, for example, and a price ceiling has a greater immediate effect on supply.

The most basic result of economic theory is that competitive markets offer the best allocation of resources possible in the absence of distortions. The optimality of competitive market prices extends to the dynamic equilibrium of a competitive market facing uncertainty and adjustment costs. However severe the premium and capacity cycle is in the insurance market, restrictions on the premiums charged cannot result in a gain in efficiency, that is, in the total surplus that agents achieve in the market. The goal of public policy should not be to stabilize the premiums in the market; it should be to maximize the gains that consumers realize from the market. This goal cannot be met with premium regulation.

D. Repeal of McCarran-Ferguson

The McCarran-Ferguson Act, exempting the insurance industry from Federal antitrust law, was passed in 1945. The industry's pricing conventions and methods are protected by the statute. These include the sharing of claims data and the establishment of premium rate levels that have historically served as mandatory rates, or as "benchmark" premium rates for insurers.

A conclusion of this Article is that the insurance crisis cannot be attributed to a lack of competition in the industry. The benefits of repealing the McCarran-Ferguson Act in improving the market stability would be modest.

124. Ch. 20, § 2, 59 Stat. 34 (codified as amended at 15 U.S.C. §§ 1011-1017 (1982 & Supp. III 1985)). The act was passed in response to a U.S. Supreme Court decision, United States v. South-Eastern Underwriters Ass'n, 322 U.S. 533 (1944), that insurance transactions that cross state lines are interstate commerce, and therefore subject to federal antitrust law.
126. One might suppose that the main effect of a repeal of McCarran-Ferguson would be to increase competition by inhibiting the state support of premium floors, which have been in place to some degree (for example, in the regulatory requirement of prior approval for rate cuts below bureau rates) in various states and at different times. This would be a potential efficiency-enhancing outcome of the repeal, since some studies have shown that rates tend to be higher in regulated states. For a summary, see TORT POLICY UPDATE, supra note 1, at app. 12-13.

However, the effect of repeal on collective rate-setting activities related to state regulation would

496
for statutory exemption from antitrust laws, this conclusion motivates consideration of the potential costs of repealing the exemption. A major potential effect of a repeal of McCarran-Ferguson would be the inhibition of claims information sharing by firms in the industry. Information about losses in hundreds of categories of risks is currently compiled and distributed in the industry largely through the industry’s Insurance Services Office.\textsuperscript{187} The bureau rates are determined by marking up the realized claims in each category. The collective organization and dissemination of information about costs could conceivably be seen to run afoul of antitrust laws, if not shielded by the Act.\textsuperscript{128}

Collective investment in information is efficient in the insurance industry. If investment in information were left entirely up to individual firms, with sharing prohibited, inefficiencies of several kinds would arise. The first set of inefficiencies would be in the pattern of information use by firms. Information about what premiums rates should be set is one of the main inputs into the production of insurance. Pooling and sharing the data about different loss experiences is efficient because information is a public good; its use by one firm does not preclude its use by others. Given the expenditures on information by firms in the market, premiums will more accurately reflect the aggregate information if all firms share in this information, and sharing avoids replication of informational expenditures by firms.\textsuperscript{129}

Second, any decrease in the ability of firms to share claims data is bound to result in an equilibrium in which firms use less information in setting premiums. Insofar as this would exacerbate the adverse selection problems that result from the asymmetry in information between firms and buyers of insurance, it would make the insurance market equilibrium more cyclical, and availability problems in general would worsen.\textsuperscript{130}

Finally, since the cost of information is a fixed cost in production, any increase in this cost would harm smaller firms proportionately more than larger firms. This would raise the minimum efficient size of firms in the likely be minimal because of the overlap between the McCarran-Ferguson exemption and the "state action doctrine," another antitrust exemption dating from 1943. See Parker v. Brown, 317 U.S. 341 (1943). The state action doctrine holds that the federal antitrust laws do not prohibit anticompetitive conduct properly attributable to the state rather than to private parties. Tort Policy Update, supra note 1 at app. 12. But see Lefkin, supra note 125, 422-24.

127. See id.
128. See Bill Would Put Insurers under Antitrust Law, supra note 5; Stewart, supra note 2.
129. See Lefkin, supra note 125, 420-22.
130. In addition, greater variation in the information sets across firms in the market would actually add another adverse selection effect in the form of the "winner's curse": the gain in a firm's share of the market for a particular risk would imply that the firm's information was overly optimistic, relative to the aggregate information of firms in the market. This effect would discourage low pricing and availability as well.
market, which would increase the market concentration. A less competitive market structure would result.\footnote{The most recent proposal to repeal McCarran-Ferguson, S. 1299, 100th Cong., 1st Sess. (1987) (sponsored by Sens. Biden, Kennedy, Metzenbaum, and Simon), departs from earlier repeal efforts in that it explicitly allows firms to share historical data on claims and reserves. Such a guarantee could be an important part of any repeal of the antitrust exemption of the insurance industry. For a discussion of the support for this bill, see Bill Would Put Insurers under Antitrust Law, supra note 5.}

Conclusion

The theory of insurance market dynamics outlined in this Article shows that the pattern of periodic crises of rising premiums and tight capacity, following periods of stable premiums and low returns, is inevitable in a competitive insurance market that operates in an uncertain environment. There are six specific, positive implications of the theory.

First, the crisis in premiums, availability, and adequacy in 1984 followed a period of stable premiums and declining returns, and resulted not only from an increase in expected costs and uncertainty in that year, but also from the cumulative losses of previous years. Second, in general a crisis will be characterized by an increase in premiums that is greater—possibly much greater—than could be “justified” by any increase in expected claims. This increase and the consequent increase in profits is consistent with a competitive market. The market wide increase in profits does not imply collusion.

A third implication is that the more uncertain the claims in an insurance market, the greater the amplitude of the cycle. The increasing socio-legal and technological uncertainties faced by the insurance market over the last decade could account for its increased cyclicality. In terms of the cross-sectional pattern of pricing in an insurance market, those lines with greatest uncertainty will experience the most severe cycles. Several features of competitive insurance markets combine to explain the suddenness and severity of crises. The most basic is the shape of the short run supply curve, that is, the sharp decrease in supply elasticity at moderate quantities of coverage. Other features are the combination of the multiproduct nature of the insurance market, the mobility of capital across lines and the variation across lines in the degree of uncertainty. The most uncertain lines will bear the brunt of shocks to the capacity of the entire market, absorbing and releasing capacity over the cycle. The adverse selection effect of imperfect risk pooling, that is, the asymmetry between insurers and insurees in the information about risks and contingent losses adds to the vulnerability of insurance markets to cycles. To the extent that this adverse selection is greater because of expanding liability standards in tort
law, increased adverse selection has contributed to increased instability or cyclicality of the insurance market.

The fourth potentially destabilizing feature of uncertain insurance lines is the effect of solvency regulation based on restrictions on the revenue to surplus ratio. This form of regulatory constraint magnifies the amplitude of the cycle since its shadow cost is pro-cyclical. It also contributes to the catastrophic timing of crises by sharpening the drop in elasticity of supply at the quantity of coverage where the capacity constraint becomes binding, even to the extent of making the short run supply curve backwards-bending.

A fifth implication of the analysis is that to the extent that the "trapped equity effect" of stockholder dividend tax liabilities is a major adjustment cost in the market or the double taxation of income a cost of maintaining surplus in a stock company, the insurance cycle is the manifestation of a tax distortion. The sixth implication is that in linking the features of the insurance crisis to the developments in tort law, the distinction between uncertainty in liability standards (the frequency of claims) and uncertainty in liability awards is important. The limits on coverage offered on policies can explained by uncertainty in the size of tort awards. The withdrawal of insurers from particular lines of insurance can be explained only with uncertainty in liability standards.

The most important policy conclusion is that regulation of premiums is likely to exacerbate problems of availability and pricing rather than solve them. However unstable the liability insurance market equilibrium may appear, it cannot be improved with premium restrictions such as those recommended by the Cuomo Commission and implemented in a number of states. In addition, the performance of the liability insurance market is entirely consistent with its competitive structure. While this is not a justification for the industry's statutory exemption from federal antitrust laws, it does suggest that the benefits of a repeal of this exemption would probably be small. The potential costs of a repeal, which include disruption of the industry's practice of sharing information on claims, should therefore be minimized.

132. See Priest, supra note 9.