JOURNAL OF FINANCIAL INTERMEDIATION 3, 379–415 (1994)

The Dynamics of Competitive Insurance Markets*

RALPH A. WINTER

Institute for Policy Analysis, University of Toronto, Toronto, Canada M5S 1A1

Received May 8, 1990

According to conventional theory, insurance premiums should be informationally efficient predictors of the present value of policy claims and expenses. This paper develops an alternative theory of insurance market dynamics based on two assumptions. First, insured risks are dependent. Under this assumption, insurers' net worth determines the market capacity since it is necessary to back the contractual promises to pay claims. Second, in raising net worth, external equity is more costly than internal equity. The theory explains the variation in premiums and insurance contracts over the "insurance cycle" and is supported by tests on postwar data. *Journal of Economic Literature* Classification Numbers: G1, G22.

1. INTRODUCTION

The conventional economic theory of insurance pricing is at odds with the dynamics of actual markets. The theory is implicit in most economic discussions of insurance pricing and is developed explicitly in the literature (Krauss and Ross, 1982; Hill, 1979; Fairley, 1979). It states that in competitive insurance markets, premiums equal the present value of expected policy claims and other expenses. Like the prices in any security market, premiums are the informationally efficient predictors of discounted cash flows. Insurance *cycles*, or persistent fluctuations in premiums relative to discounted claims, are impossible in this theory.

* I am grateful for comments and suggestions from numerous colleagues, two referees, and seminar participants at MIT, Northwestern, NBER, Queen's, Toronto, and Yale. The editors provided extraordinary guidance during the revision of this paper, and Joseph Stiglitz offered comments on a related paper that were especially helpful here. Finally, I am indebted to George Priest for his encouragement and contributions to the ideas in this paper. This work was supported by an Olin Fellowship in the Civil Liability Program at Yale Law School and a grant from the Social Sciences and Humanities Research Council of Canada.

1042-9573/94 $6.00

In reality, the supply in property-liability insurance markets appears to fluctuate, especially in liability lines. "Soft markets" of stable premiums and low returns to insurers are followed by "tight markets" of rapidly rising premiums. In the 1970s, a soft market was followed by tight supply in medical malpractice and products liability insurance. In the 1980s, a soft market was followed by the widespread "insurance crisis" of 1984–1986.[1] In both of these episodes, many premiums rose by hundreds of percent, tripling for General Liability lines between 1984 and 1986 (see Best (1989), p. 87). In mid-1987, the market turned around again, with premiums falling by as much as 40% in some lines.

This paper offers an alternative theory of pricing in competitive insurance markets and tests the theory with postwar data in the U.S. property-liability insurance markets. The theory attributes jumps in premiums to the accumulation of losses, not just to contemporaneous changes in expected claims. It links the sudden crisis in the supply of liability insurance in the 1980s to the expansion of liability and awards in the U.S. tort system (Priest, 1987). The theory explains two additional features of the insurance cycle. In tight markets, prices are not only high but *nonlinear*. Price becomes a steeper function of coverage; i.e., smaller amounts of coverage are available at any given premium.[2] In many lines transactions dry up altogether in a tight market.[3] A theory of the cycle must explain this absence of gains to trade in particular lines during tight markets. The three features of a tight market are captured in the common description of the 1984–1986 experience as a "crisis of availability, adequacy, and affordability."

The model below is driven by two assumptions. First, insured risks are dependent. Average claims cannot be predicted with certainty because of aggregate uncertainty or common factors. An important common factor is

[1] The crisis of 1984–1986 was a remarkable disruption in the performance of the liability insurance market. See Tort Policy Working Group (1986, 1987).

[2] Over the 1984–1986 period, the tightening of limits was dramatic in liability insurance lines. For example, major pharmaceutical companies typically purchased over 200 million dollars in the "excess coverage" line in early 1984; by 1986 excess coverage of more than 10 million was difficult to obtain. With the turning point in mid-1987, coverage limits increased up to 500% compared to the previous year. For a description of the turnaround, see *Business Insurance*, p. 25, July 4, 1988.

[3] In 1984–1986, insurance transactions disappeared almost completely in lines such as municipality liability, directors' and officers' liability, and day-care liability insurance. This phenomenon is usually described as saying that "insurance was unavailable at any price," but of course this description is incorrect. There is always a price at which supply will be positive. In instances where individual's demand for insurance was inelastic because of legal requirements to have insurance coverage, policies were offered at extremely high premiums. Firms removing asbestos from buildings paid premiums that were in some cases more than 90% of the coverage limits!

tort law, which determines losses in liability insurance lines. Dependence of risks means that insurers with limited liability must maintain enough equity or net worth to give credibility to their promises to pay claims. Equity becomes a measure of capacity in the market. The second assumption is that in raising net worth for insurers, internal capital is less costly than external equity. This makes equity a state variable in the model. Tight markets correspond to low levels of equity and persist because insurers prefer to wait out the rapid accumulation of retained earnings rather than resort to costly external capital. Soft markets persist because insurers recognize the possibility that the excess stock of internal equity will be needed in the future and are therefore reluctant to distribute all excess cash to shareholders. These effects are developed in a recursive equilibrium framework (Stokey and Lucas, 1989) in which the three features of the insurance cycle are traced to basic market conditions.

The model is essentially a version of the "finance-constrained firm" approach to the theory of business cycles (Greenwald and Stiglitz, (1993). The insurance market is a persuasive context for this theory because other constraints on supply, such as physical capital, are unimportant. Grøn (1989, 1994) independently develops the implications of the capacity-constraint model and offers an impressive array of tests. This paper complements Grøn's in developing the theoretical foundations. The econometric evidence below focuses on the most basic implication of the model.[4] The paper contributes also to the literature on optimal insurance contracts by offering a new explanation of limits on insurance coverage.

Section 2 of this article contains the basic model of premium dynamics in a simple supply–demand model. Section 3 extends the assumptions to a multiple-line insurance market. This section explains the disappearance of gains to trade for some lines of insurance and the cross-sectional puzzle of why only some lines (commercial casualty lines) are much affected by tightening capacity. Section 4 extends the assumptions to explain the pattern of nonlinear pricing over the cycle. Section 5 tests the model with

[4] Other papers have attributed the insurance cycle to psychological effects and capital mobility (Stewart, 1984); to naive rate-making extrapolations of past claims data (Venezian, 1985); to informational asymmetries and cream-skimming (Nye and Hofflander, 1987); to contracting and informational and regulatory rigidities, such as lags in data collection, in regulatory approvals, or in contract adjustments at policy renewal (Cummins and Outrevile, 1987); or to continuing attempts by the market to clear in a partial adjustment model in which firms are constrained in the short run by the capacity of production channels, in particular distribution channels (Doherty and Bang, 1988). Berger (1988) offers a simple "reduced-form" model of price and capacity fluctuations in insurance markets but does not develop it from first principles. Danzon and Harrington (1992) offer an excellent review of the literature on insurance cycles and other aspects of liability insurance markets. In a previous paper (Winter, 1988), I offered a nontechnical discussion of the capacity-constraint hypothesis as a basis for evaluating policy responses to the instability in insurance supply.

postwar time-series data and against the stylized facts of the 1984–1986 experience. The concluding section summarizes the results.

2. PREMIUM DYNAMICS

The dominant feature of modern property-liability insurance markets is the uncertainty faced by insurers in predicting claims. If risks were independent, the law of large numbers would guarantee that average claims were accurately predictable. But independence fails because of common factors in the distribution of claims. The most important common factors are the trends in tort law, both in terms of the negligence standards and the damages awarded by courts, especially in jury trials (Priest, 1987; Trebilcock, 1987); the future legal liability of insurers (Romano, 1989); and natural or technological factors, e.g., the frequency of child abuse in day-care centers or the future side effects of drugs.[5]

Given uncertainty in average claims, the net worth of an insurer with limited liability measures the insurer's *capacity*. An insurer is constrained in the amount of insurance it can write at moderate premiums by the credibility of its promise to cover claims in adverse states of nature. Alternatively, regulatory constraints on solvency may determine the quantity of coverage that a firm can offer given its net worth.[6]

Maintaining net worth is costly for insurers. An extra level of taxation is imposed on funds invested through the insurance corporation rather than by shareholders directly.[7] In addition, agency costs are incurred as a result of conflicting interests of management, shareholders, and policyholders.

A second key feature of the property-liability insurance market is that internal equity is less costly than external equity. By this I mean that a cost is incurred in the "round trip" of paying out a substantial amount of

[5] Liability insurance involves risks that are not resolved until many years after the payment of the premium. The buyer of a one-year standard liability insurance policy is covered for liability for accidents which are caused during the policy year, even if the losses are not manifest until much later. Thus forecasting the average losses on a group of policies requires a prediction about how the common factors, such as the legal standard of liability, will evolve.

[6] Winter (1991a) considers in detail the implications of solvency regulation for the insurance cycle. For present purposes, we assume simply that regulation restricts the probability of bankruptcy to zero.

[7] The opportunity cost of double taxation, including the foregone return from holding tax-exempt securities, is demonstrated by the fact that in 1986, 30% of total investment income of property-liability insurance companies was from tax-exempt bonds (calculated from Best (1987, p. 54)). This income would not be taxed if held directly by shareholders but *is* taxed when distributed as dividends from the corporations.

retained earnings in cash to shareholders and then immediately raising the same amount through the issuance of equity. For empirical evidence on the cost advantage of internal capital and a general reluctance of corporations to resort to external equity see Asquith and Mullins (1986) and Auerbach (1983). One explanation for this difference in costs is the superiority of information of firm managers over outside suppliers of equity (Stiglitz, 1982; Myers and Majluf, 1984). The argument is that issuing external equity signals that expected profits are relatively low. A second advantage to internal equity derives from the "trapped-equity effect" of dividend taxation.[8] Finally, an industry-specific regulation discourages the exit of equity from the market: insurers are constrained in most states against distributing more than 10% of net worth to shareholders in any year.[9]

2.1. *The Model*

The *ex ante* distribution of the payout on a typical insurance policy has a spike at zero and a gap in its support between zero and some minimum loss. Without loss of generality, one can describe such a distribution as arising from a two-stage lottery: "loss or no loss," and "size of loss." For example, in a liability insurance policy, the risk consists of the event of liability and then the damages. The implications for the insurance market of dependence in each of these stages differ, and it is instructive to consider the two types of dependence separately. This section considers dependence only in the events of losses.

The model is in discrete time. I will describe the events and transactions of the model starting at the end of each period and work backward. A large number of economic agents each face the risk of losing a dollar at the end of the period. The probability of a loss is itself unknown. That is, each risk takes the form of a two-stage, compound lottery. In the first stage of the lottery, a random probability is drawn; in the second stage, for each individual the loss of a dollar occurs with the realized probability.

[8] Once equity is raised, it cannot be transferred back to shareholders without incurring individual income tax. This barrier to the exit of capital is mitigated by the possibility of share repurchases (Bagwell and Shoven, 1989), but the tax code prevents tax avoidance through the complete reliance on repurchases to distribute cash.

[9] For example, see New York State insurance statutes, section 4105, and Massachusetts statutes, section 72. New York has an additional constraint that dividends not exceed net investment income (with adjustments). Firms selling insurance in any state are constrained by that state's regulations in all of their business. The purpose of these statutory constraints is to guard against the agency problem of excess dividends at the expense of senior stakeholders, in this case past purchasers of insurance policies. An examination of the data shows that this constraint was not binding for all firms in the "soft market" of the early 1980s, but probably was for some. For example, Aetna paid out 9.95% of its surplus as dividends in 1983.

The realization of the random probability, p_t, is identical across individuals, whereas the events of losses conditional upon this probability are independent across individuals. The probabilities $\{p_t\}$ are taken to be independent and, for simplicity, identically distributed over time according to a cumulative distribution function F with density f. The support of F is $[0, p_H]$.

Prior to the realization of losses in each period a market opens for insurance, which is defined as dollars paid contingent upon the event of a loss to an individual. This is the only contingent market assumed to exist; agents in this model cannot buy insurance for future periods. One interpretation is that a new generation of consumers enters in each period. A "period" should be interpreted as the maturity of a contract; I am ignoring the fact that contract periods are overlapping. Assuming that agents are risk-averse, expected utility maximizers, the demand for insurance depends upon the risks only through $\bar{p}$, the mean of p_t under F or the *ex ante* probability of an accident for any individual.[10] We take the demand curve to be downward-sloping.[11]

The supply side is competitive and in the usual way can be treated as an aggregate, price-taking firm. Insurers have limited liability. Having entered the period with surplus or net worth X_t, the aggregate supplier can adjust its surplus prior to the opening of the insurance market. It adds to its surplus by contributing equity, e_t, or reduces its surplus by a payout, d_t, to shareholders. The adjusted surplus is denoted by $S_t = X_t + e_t - d_t$.

The cost advantage of retained earnings is introduced in the simplest way possible. A cost (e.g., a transactions or signaling cost) of k_1 dollars is incurred with the entry of equity, and a cost k_2 is incurred with exit. Both costs are paid directly by shareholders. A special case is the trapped equity model of dividend taxation, in which $k_1 = 0$ and $k_2 = t$, where t is a tax rate.[12] Consumers can observe the adjusted surplus when the market for insurance opens. If this surplus plus premium revenues were insufficient to cover the insurer's liabilities under some realizations of p_t, then the insurance claims would be met only partially.

I assume in this section that insurance is issued only up to the limit of no bankruptcy. If consumers are infinitely risk averse this is an optimal

[10] The independence axiom implies that uncertainty in the probability of a binary event is irrelevant to an expected utility maximizer.

[11] This assumption holds except for extreme values of risk-aversion parameters (Hoy and Robson, 1981; Briggs *et al.*, 1989).

[12] I do not incorporate explicit regulatory constraints on the flow of dividends, but the extension is straightforward—conceptually if not technically. The assumption of linear adjustment costs yields some unrealistic implications (e.g., a bang-bang adjustment of capital), but allows the simplest model of premium and contract dynamics. Furthermore, linearity of adjustment costs is a much closer approximation to reality when capacity is financial capital, or equity, than physical capital.

Period t	Period t+1
• aggregate firm enters with surplus X_t	$X_{t+1} = [S_t + (P_t - p_t)Q_t] \cdot (1 + r - c)$
• chooses d_t or e_t $S_t \equiv X_t + e_t - d_t$	
• insurance market opens P_t and Q_t determined	
• p_t realized and claims paid	

FIG. 1. Timing of the model.

policy under the assumptions already specified. To the extent that they are not, a bankruptcy cost is implicit in the assumption.[13] The market price and quantity of insurance in period t are denoted by P_t and Q_t.

Finally, the surplus or net worth of an insurer at the end of the period, which equals the adjusted surplus plus underwriting profit, accrues at a rate of $r - c$ to become the unadjusted surplus at the beginning of the next period. Here, r is the economy's (riskless) interest rate after personal taxes, and c is the flow cost of maintaining equity which was discussed at the beginning of this section. Thus,

$$X_{t+1} = [S_t + (P_t - p_t)Q_t](1 + r - c). \tag{1}$$

The timing of the model is summarized in Fig. 1. The firm's objective in its decisions on the quantity of insurance offered and dividends and equity issued is to maximize the present value of net cash flows to shareholders. These cash flows equal dividends net of adjustment costs, minus equity issued. The risk-neutrality of shareholders reflects an assumption that the exogenous uncertainty in accident frequency is independent of aggregate wealth in the economy. The expected present value of dividends per dollar of surplus in the market is the market-to-book or Tobin's q ratio, denoted by q.[14]

[13] The assumption of bankruptcy costs is relaxed in Section 3.

[14] Mutuals and other cooperative insurers do not appear explicitly in the model. Mutual insurance is an inherently inferior means of risk-bearing when risks are dependent since cooperative insurance offers no protection against common factors; but mutuals have offsetting advantages in that policyholder dividends are not taxed and the gains to mutual insurance are insensitive to asymmetry of information about aggregate uncertainty (Smith and Stutzer, 1990). If the model were extended to incorporate mutuals, both organizational forms would exist in equilibrium. The utility function in our model can be interpreted as the utility derived from a cooperative subgame in which individuals pool risks. The reinsurance market also does not appear explicitly in the model. The insurance market in this model can be interpreted as the aggregate of the direct insurance market and the reinsurance market; the model is too simple to explain the separate existence of reinsurance. In reality, capacity was even tighter in the world reinsurance market than in the U.S. direct insurance market in 1985 (see Section 5, this paper, and Berger *et al.* (1992)).

2.2. *The Equilibrium*

I exploit the Markov structure of the model through the concept of a *recursive competitive equilibrium* (Stokey and Lucas, 1989). A single state variable, X, the surplus, describes the market at the beginning of any period. A central condition of equilibrium is that firms—equivalently, an aggregate firm—maximize profits, taking as parametric both the current prices q, P and the expectation of the market price of equity in each "state of the world" (realization of p) *next* period, given today's state X. This expectation is denoted $q^+(X, p)$. The seven-tuple $[d^*(X), e^*(X), Q^*(X); P^*(X), q^*(X); q^+(X, p); \tilde{X}^*(X, p)]$ is an equilibrium if it satisfies the following five conditions.

(i) $\tilde{X}(X, p) = (1 + r - c)(X + e(X) - d(X) + [P(X) - p]Q(X))$

(ii) $d(X)$, $e(X)$, and $Q(X)$ solve, at each X:

$$\max_{d,e,Q} E[q^+(X, p) \cdot (1 + r - c)(X + e - d + [P(X) - p]Q)]/(1 + r) + (1 - k_2)d - (1 + k_1)e \tag{2}$$

$$\text{subject to} \quad X + e - d + P(X)Q \geq p_H Q \tag{3}$$

(iii) $Q(X) = D(P(X))$

(iv) $q(X)X = (1 - k_2)d(X) - (1 + k_1)e(X) + E[q^+(X, p)\tilde{X}(X, p)]/(1 + r)$

(v) $q^+(X, p) = q(\tilde{X}(X, p))$.

Condition (i) is a transition equation for the surplus X. Condition (ii) states that the aggregate competitive firm maximizes the expected present value of its next-period value, plus current dividends minus current equity issued. Note that $q^+(X, p)$ is taken as fixed in this maximization problem by the price-taking firm. Condition (iii) is a market-clearing condition; (iv) is a recursive valuation equation, and (v) is a condition on the rationality of expectations in the recursive model: The price of equity anticipated for next period at each realization of p must be the price actually assigned to next period's equity by the stationary, equilibrium price function $q(X)$.

A standard reformulation of the recursive equilibrium starts with the supposition that the market anticipated a functional relationship $\bar{q}(\tilde{X})$ between the price of equity and capacity in the next period, and then examines the consequence for equilibrium prices and quantity in the current period, in particular $q(X)$. This defines an operator ψ via $q \equiv \psi(\bar{q})$. The concept of agent rationality embodied in our equilibrium definition can be broken into two parts. First, the competitive firm expects a particular price of equity, $q^+(X, p)$, in each state of the world (realization of p) next period and takes this as fixed. This expectation must be self-realizing, or rational. Second, the value function q^* must be self-replicating, i.e., a fixed point of the operator ψ. It is useful to isolate the first of these two rationality conditions by defining a *conditional rational expectations equilibrium* (CREE).

Given an arbitrary, increasing function $\bar{q}(\tilde{X})$, the conditional rational expectations equilibrium is a solution to the equilibrium conditions (i) through (iv) and

$$(\text{v}') \quad q^+(X, p) = \bar{q}(\tilde{X}(X, p)).$$

This equilibrium concept differs from the recursive equilibrium insofar as $\bar{q}$ appears on the right-hand side of (v′) instead of q. In the CREE concept, we are not asking that the market's anticipated value function $\bar{q}$ be generated as the value function in the current period, only that the expectations $q^+(X, p)$ be consistent with $\bar{q}$ and the transition function $\tilde{X}(X, p)$. The CREE describes what the equilibrium would be if equity-holders actually *were* to be compensated with $\bar{q}$ next period. Imposing the additional condition on a CREE that $q = \bar{q}$, i.e., that $\bar{q}$ be self-realizing, yields a full recursive equilibrium.

When consumer preferences are near risk-neutrality, there are no gains to trade in this market, due to the capital market imperfections. It can be shown, however, that a CREE exists if preferences are sufficiently risk-averse and if $\bar{q} \in M$, the set of (value) functions satisfying the following restrictions: $q(X)$ is bounded by $(1 - k_2)$ and $(1 + k_1)$; $q(X)$ is nonincreasing; and $Xq(X)$ is nondecreasing.[15] Furthermore, the operator ψ preserves these properties and is therefore an operator on M. I have verified computationally that for a range of parameters in this model, there is a fixed point $q^*(X)$ of the operator ψ in M. The corresponding price function $P^*(X)$ is strictly decreasing in the computed equilibria.

The predictions that flow from this model are based on the characterization of the recursive equilibrium, which we assume to exist and satisfy $q^*(X) \in M$ and $P^*(X)$ strictly decreasing, as the computed examples do. It is useful to describe the CREE given $\bar{q}(X) \in M$ both as a basis for these predictions and to indicate the construction of the operator ψ for computing the equilibria. To describe a CREE given the state X, we start at the end of a period and consider the equilibrium price and quantity in the insurance market given the market's anticipation of $\bar{q}$ next period and the equilibrium functions $\tilde{X}(X, p)$ and $q^+(X, p)$.

This equilibrium can be described b the intersection of the exogenous demand curve with an endogenous short-run supply curve. From (ii), the supply at P is the solution to the problem

$$\max_{Q} E[q^+(X, p) \cdot (1 + r - c)(X + e - d + [P - p]Q)]/(1 + r), \quad (4)$$

[15] Details are available from the author.

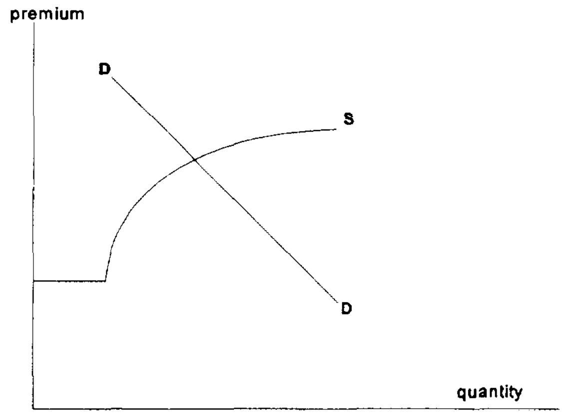

FIG. 2. Short-run equilibrium with tight capacity.

subject to the limited-liability constraint (3). Since the objective function (4) is linear in Q, the supply curve is horizontal to a limit determined by the limited-liablity constraint. The horizontal section of the supply curve is at the price, denoted by $\hat{P}(X; \tilde{q})$, at which the derivative of the objective function in (4) is zero. Differentiating (4) and substituting (v') shows that $\hat{P}(X, \tilde{q})$ is the solution, in P, to

$$E(P - p)\tilde{q}(\tilde{X}(X, p)) = 0. \tag{5}$$

Solving (5) yields

$$\hat{P}(X; \tilde{q}) = \bar{p} + \frac{\operatorname{cov}(\tilde{q}(\tilde{X}(X, p)), p)}{E\tilde{q}(\tilde{X}(X, p))}. \tag{6}$$

Next, define the price $P_c(X)$ as the solution to market clearing, (iii), and the limited-liability constraint (3) as an equality; this is the price that will obtain when the limited-liability constraint, or capacity constraint, is binding. The equilibrium price at X, given $\tilde{q}(\cdot)$, is the maximum of $\hat{P}(X; \tilde{q})$ and $P_c(X)$. Short-run equilibria are depicted in Fig. 2, for the case of a binding limited-liability constraint or "tight market," and Fig. 3 for the case of a nonbinding constraint or soft market.

Having described the equilibrium in the insurance market of the current period, we can move backward one step to describe the equilibrium adjustment of equity. The value function in the current period, if there were *no* adjustment, would be the discounted total expected value of equity next period, per unit of current equity, or from (i), (iv), and (v'):

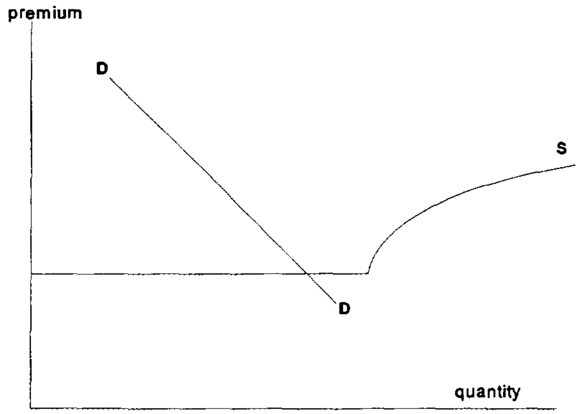

FIG. 3. Short-run equilibrium with excess capacity.

$$\hat{q}(X) \equiv \frac{1}{X}\frac{1}{(1+r)} E[\tilde{q}((1+r-c)[X+[P(X)-p]Q(X)]) \cdot (1+r-c)(X+[P(X)-p]Q(X))]. \tag{7}$$

One can show analytically that this function in nonincreasing in X, for $\tilde{q} \in M$. The effect of entry and exit of equity in the current period is to bound the value function $q(X)$ in the current period between $(1-k_2)$ and $(1+k_1)$. Specifically, define S_M as the minimum value of X for which $\hat{q}(X) = (1-k_2)$ and S_m as the maximum point at which $\hat{q}(X) = (1+k_1)$.[16] Define $d(X)$ as $\max(0, X - S_m)$ and $e(X) = \max(0, S_M - X)$. The functions $d(X)$ and $e(X)$ describe the optimal adjustment of equity. For example, if $X < S_m$ then $q(X) = (1+k_1)$ and the price-taking firm is indifferent among all nonnegative injections of equity into the market including, in particular, the entry of $S_m - X$.

Given $\tilde{q}$ the value of the operator ψ is thus described by

$$q(X) = \frac{E[\tilde{q}(\tilde{X}(X, p)) \cdot \tilde{X}(X, p)]}{X(1+r)}, \tag{8}$$

truncated at $(1-k_2)$ and $(1+k_1)$. A fixed point of this operator together with the associated price function and limits on equity constitute a recursive competitive equilibrium of the model.

[16] In computing the $\psi(\tilde{q})$, one artificially constrains the equity to be adjusted within some bounds. A fixed point q^* is an equilibrium if the truncation points of q^* lie within these artificial bounds.

Note that in the case where the limited-liability constraint is not binding, (i), (5), and (8) imply that $q(X) = [(1 + r - c)/(1 + r)] \cdot E[\tilde{q}(\tilde{X}(X, p))$. Thus, in this case a competitive firm could earn its market value simply by investing in the riskless asset.

2.3. *Properties of the Equilibrium*

The model offers a number of implications for competitive insurance markets. The first is a basic point about the equilibrium premium and the allocation of risk-bearing in the market.

PROPOSITION 1. *For an equilibrium satisfying $q^* \in M$, the equilibrium premium $P^*(X)$ strictly exceeds the mean $\bar{p}$ at any state X, even when the capacity constraint is nonbinding. The equilibrium insurance coverage involves less than full insurance coverage, for finite risk aversion.*

This proposition follows directly from Eq. (6). Because $q^*(X)$ is decreasing in X, and X is decreasing in the realized p, the covariance term in Eq. (6) is positive. Even when the capacity constraint is not binding, the fundamental result of conventional insurance theory—that risk-neutral agents (shareholders in this model) should bear all the risk—fails to hold. The objective of maximizing expected value mimics risk aversion because of the correlation of losses across firms and the monotonicity of q; the event of a high loss for the aggregate firm corresponds to a high marginal opportunity cost of additional claims payment.

The most important testable implication of this model is a failure of the present value relationship. The computed $P^*(X)$ is, as described above, decreasing in X. The generalization to nonstationary risks is straightforward: the error in the premium as a predictor of the expected present value of claims is negatively related to the current surplus. Premiums are not informationally efficient predictors of future claims even with rational expectations.

The model can explain the pattern of relatively sharp changes in industry prices. For contrast, consider a Marshallian model of a conventional product market in which the short-run supply curve is derived from a neoclassical technology with a given amount of some fixed factor. In this model, the equilibrium price is typically convex in the capacity or amount of the fixed factor. That is, the marginal impact of declining capacity on price is initially small and then increases. In our model, however, the next proposition shows that the equilibrium price is typically *concave* in capacity, where the limited-liability constraint is binding. Let the profit function be denoted by $\pi(P, p) \equiv (P - p)D(P)$; we make the regularity assumption that profit is concave.

PROPOSITION 2. *Suppose that $D'(P) < 0$ (risk aversion is finite) and that $\pi(P, p)$ is concave in P for $p = p_H$. Then over any interval where the*

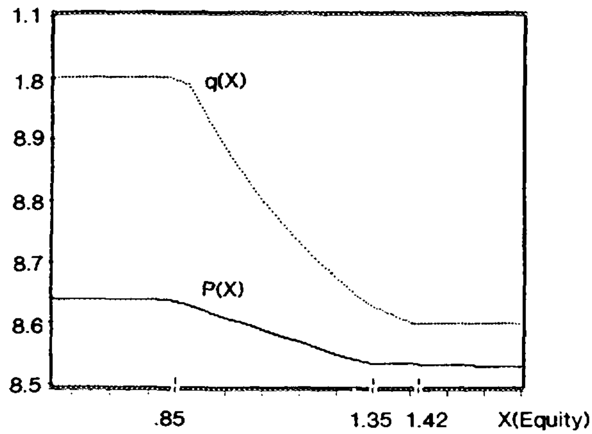

FIG. 4. Computed equilibrium value function and price function.

limited-liability constraint is binding, the equilibrium price function $P^*(X)$ is concave.

Proof. When the limited-liability constraint (3) is binding, $-X = \pi(P, p_H)$ or $P^*(X) = \pi^{-1}(-X, p_H)$. This is concave when π is concave.

The marginal impact of declining capacity, $|dP/dX|$, is therefore highest when the limited-liability constraint is just binding. An additional property of the equilibrium, revealed in simulations, is that the price $\hat{P}(X; \bar{q})$, established when the limited-liability constraint is nonbinding, is relatively insensitive to changes in capacity. This is illustrated in Fig. 4. Together with this property, Proposition 3 implies that the impact of capacity shocks can be more sudden in insurance markets than in other markets. Capacity contraints in the insurance market are relatively sudden, according to the model, and tend to follow periods of stable premiums and low returns. Insurance "crises" result from cumulative losses rather than contemporaneous shocks alone. When the market is in a period of excess capacity (Fig. 3), losses will not initially affect the premium by much (since $\hat{P}$ is nearly constant). Then, when capacity is sufficiently depleted so that the limited-liability constraint becomes binding, price rises quickly.[17]

The basic point of the model, that premiums will fluctuate with capacity and not just with changes in the distribution of risks, is made most directly

[17] This argument assumes a very sharp limited-liability constraint. In practice, some smoothness in the impact of the capacity constraint will result from heterogeneity among firms. However, this is true in a conventional market as well. The emphasis here is on the unique aspects of insurance supply under limited capacity.

under our assumption of stationary risks. It is nonetheless useful to consider the impact of an unanticipated, once-and-for-all increase in uncertainty. Consider, for example, a single tort decision that raises the frequency of losses under the "worst-case" scenario. The short-run impact is an increase in premiums *greater* than the increase in expected costs, whenever the market is capacity-constrained:

PROPOSITION 3. *Suppose that the distribution F has an atom at* p_H *and that the limited liability constraint is binding in equilibrium. With changes in* p_H, $dP/dp_H > d\bar{p}/dp_H$.

Proof. $dP(X)/dp_H = 1$ by Eq. (3), whereas $d\bar{p}/dp_H = \text{prob}(p_H) < 1$.

Unless the elasticity of demand for insurance is high, expected profits will also rise with the increase in p_H. Thus, a competitive insurance market may respond more to changes in the high end of the claim distribution than to the mean of the distribution. For example, a relatively small number of court decisions can have a dramatic effect. A consequence of such a rise in premiums above costs will be an increase in profits, providing that elasticity of demand is sufficiently low. Surprisingly, a *competitive* insurance market may appear to exaggerate tort awards so as to justify premiums increases that add to profits, a criticism that was leveled against the industry during the mid-1980s crisis.

While the simple model of this section is intended to provide qualitative predictions rather than an empirically realistic description of the market, it is nonetheless important to ask if it can explain the amplitude of the insurance cycle. It cannot. Figure 4 depicts the equilibrium function $P^*(X)$ and $q^*(X)$ for the following illustrative set of parameters: rectangular demand at a quantity = 5; $c = 0.05$; $k_1 = 0$; $k_2 = 0.40$; $r = 0.3$ (e.g., a three-year maturity with an annual interest rate of about 0.09); and p uniformly distributed on [0.2, 0.8]. For these parameters, the minimum and maximum capacities are 0.85 and 1.42, and the premium varies between 0.53 and 0.64. Only a 20% range in premium fluctuations can be explained in the simple model even for extreme parameters. The extension in the next section of this paper has more power in explaining the magnitude of insurance crises.[18]

[18] Several qualitative implications of the simple model are clearly inconsistent with empirical reality. The first implication is that dividends or equity are paid only periodically and never simultaneously. The second is the prediction that the premium will follow only a first-order process, instead of a cycle of at least second order. The actual process is not first order (Section 5). The culprit here is the assumption that contracts offer coverage for only one period, i.e., are not overlapping. In the actual market, the liability incurred by an insurer entering a liability insurance policy extends for many years; policy periods are overlapping.

3. MULTIPLE LINES

With a single line of insurance as in the previous section, there are gains to trade in every period if there are any at all. A multiple-line extension of the model allows an explanation of the cross-sectional impact of a tighter capacity constraint—why only some insurance lines are much affected by tight capacity and why some dry up altogether in tight markets. The essential result of this section is a second-order comparative static proposition: The greater the uncertainty of a particular line, the more sensitive it is to fluctuations in total industry capacity. The most uncertain lines bear the brunt of fluctuations in aggregate capacity.

I consider the short-run equilibrium in one period, assuming that the dynamic equilibrium exists and yields a function q that is monotonic in total capacity. The key assumptions are first, that capacity, while specific to the market, is completely mobile across lines, and second, that lines differ in uncertainty. The introduction of multiple lines leads to two issues: the differential effects of capacity shocks across lines, and the pooling of risks across lines. To focus on the first, I assume that the source of dependence among all risks is a single common factor. Different lines have different factor loadings, or exposures to the factor.[19] In some "baseline" 0, the frequency of claims is given by p, with mean $\bar{p}$ and maximum p_{H}, as before. In line i, the realized frequency is given by $p_i \equiv a_i p + (1 - a_i)\bar{p}$ for some a_i. Thus the lines differ by mean-preserving spreads in the frequency of claims. The lines have the same demand curve; demand is unaffected by uncertainty in the frequency of losses.

The aggregate, price-taking[20] firm, after equity adjustment, chooses Q_i, $i = 1, 2 \ldots$, to solve

$$\max_{Q_i} E\{q_{t+1} \cdot \left[X + \sum_i (P_i - p_i)Q_i\right] \cdot (1 + r - c)]\}, \tag{9}$$

subject to

$$X + \sum_i (P_i - p_{\mathrm{H}i})Q_i \geq 0. \tag{10}$$

[19] For example, the source of dependence may be the trend in tort liability standards, with lines differing in their sensitivities to this factor. This simplifying assumption implies that the risks are perfectly correlated. The analysis is unchanged if lines have idiosyncratic factors and each line is small.

[20] The firm takes the prices P_i and the future equity price q_{t+1} as given, not recognizing their dependence upon Q_i.

LEMMA. *The maximization problem in* (9) *is equivalent to the following*:

$$\max_{Q_i, X_i} E\{q_{t+1} \cdot \left[\sum_i X_i + (P_i - p_i)Q_i\right] \cdot (1 + r - c)\} \tag{11}$$

subject to

$$\forall i, \qquad X_i + (P_i - p_{Hi})Q_i \geq 0 \tag{12}$$

and

$$\sum_i X_i = X. \tag{13}$$

This separation result allows the problem to be described in two stages. The amount of capacity allocated to each line is determined by equality of the shadow price of the capacity constraint across lines if the constraint is binding, and the equilibrium in each line, given its capacity, is determined exactly as in the single-line model. We are abstracting from gains to pooling across lines, with the assumption of only one common factor.

The basic predictions of the multiline extension are captured in two comparative-static propositions. The first is:

PROPOSITION 4. *The more uncertain lines have higher premiums*: $a_i > a_j$ *implies that* $P_i > P_j$ *whether the capacity constraint is binding or not.*

Proposition 4 shows that greater uncertainty in a particular line has an impact even during soft markets. To prove the proposition, note that where the capacity constraint is not binding, the equilibrium prices are given by Eq. (6) as before, and capacity is not dedicated to any line. In this case, the covariance in (6) is larger for the more uncertain lines, proving the lemma for this case. When the capacity constraint is binding, the shadow prices of the capacity constraints in all lines must be equal. This shadow price is given by the expected next-period value of allocating one more dollar to X_i. Since an additional dollar invested in line i would allow additional quantity of $(p_{Hi} - P_i)^{-1}$, this shadow price is

$$\lambda = E\left[q_{t+1} \cdot \frac{P_i - p_i}{p_{Hi} - P_i}\right]. \tag{14}$$

Solving (14) for P_i yields

$$P_i = \frac{1}{\lambda + Eq_{t+1}} [\lambda p_{Hi} + \bar{p}Eq_{t+1} + a_i \text{cov}(q_{t+1}, p)], \tag{15}$$

which is increasing in a_i since p_{Hi} is increasing in a_i and the covariance term is positive. This proves Proposition 4.

Under a regularity condition on demand, the capacity dedicated to line i is more sensitive to fluctuations in aggregate capacity the more uncertain is the line i:

PROPOSITION 5. *Assume that the elasticity of demand,* $|\partial \ln D(P)/\partial \ln P|$, *is increasing in* P, *and consider two lines* i *and* j, *with* $a_i > a_j$. *When the limited-liability constraint is binding and both* X_i *and* X_j *are positive, the elasticity of* X_i *with respect to* X *exceeds the elasticity of* X_j *with respect to* X.[21]

Proof. Proofs of remaining propositions are in Appendix 1.

As an illustration of Proposition 5, during the 1984–1986 crisis, automobile insurance premiums were relatively stable, in spite of the tightening capacity, because risks in these lines are relatively predictable.[22] The price given by (15) is increasing in the shadow price of capacity, λ. Suppose that the maximum demand price, the price at which demand falls to zero, is less than p_{Hi} for the most uncertain lines i. Then for these lines the supply price given by the equation may exceed the maximum demand price when λ is sufficiently high (i.e., when capacity is sufficiently scarce). In this case, the transactions in these lines will dry up altogether. This also explains the emergence of exclusionary clauses in policies renewed in 1984 and 1985.[23]

4. LOSS UNCERTAINTY

In the equilibrium of the simple model, all contracts are uniform (linear) pricing contracts: a consumer can buy as much coverage as desired at the prevailing market price. Which additional feature of the actual market conditions accounts for the limits on coverage and the dramatic drop in these limits in 1985? This aspect of the market performance is commonly discussed in terms of "rationing"—as if insurers had only so much coverage that they were willing to offer and doled it out in limited amounts. The explanation here is more precise.

[21] The monotonicity of demand elasticity, used in this proposition, is a standard assumption for general demand functions and was verified numerically for insurance demand in the cases of log utility and other constant relative risk-aversion utility functions.

[22] The net premiums written for liability insurance jumped 80% in 1985, whereas the net premiums written for total auto insurance rose 19% (Best, 1987, p. 103).

[23] For example, in day-care insurance, liability resulting from child abuse was excluded from many policies; in directors' liability insurance, liability arising from shareholder derivative suits was often excluded.

The extension to variation in the size of losses for each consumer is clearly necessary to explain the contract complexity. But even under this assumption, the prediction of traditional economic theory of insurance is that among all actuarially equivalent insurance contracts, an expected utility maximizer prefers the one with a deductible and full coverage (Arrow, 1970). It is the large losses a consumer most wants to insure. Furthermore, actuarially fair premiums per coverage are *declining* in the amount of coverage (Proposition 7 below). Yet in the 1985 liability insurance market, corporations paid premiums for "excess coverage" of high losses that were often three times basic coverage, and others chose, or were forced into, coverage limits that were a fraction of potential maximum losses.

Two explanations of coverage limits in insurance contracts have recently been offered in the literature: adverse selection (Rothschild and Stiglitz, 1976) and limited liability on the *buyers'* side of the market (Huberman *et al.*, 1983).[24] This section shows that limits can be explained by limited liability on the sellers side of the insurance market combined with dependence in the random size of a loss, as opposed to the event of a loss. For example, if the trend in tort law is the main source of uncertainty, then the model of Section 2 assumed uncertainty only in courts' standards of assigning liability; here we consider uncertainty in the size of damages.

The lottery faced by the N consumers in the market is replaced with the following. Consumers, who are assumed to be homogeneous, have a Bernoulli utility function $U(\cdot)$, initial wealth W, and with probability p face an accident or loss of wealth. The loss itself is uncertain, with a cumulative distribution (conditional upon the event of an accident) given by $F(L)$, with density $f(L)$. Let M be the maximum point of support of F. The first stage of the lottery (the event of an accident) is independent across consumers, whereas the realization of the random loss is, for those who incur it, identical. This structure allows us to isolate the implications of dependence in the size of losses.

With identical consumers, a single type of contract is offered in equilibrium, and without loss of generality each consumer purchases all insurance from a single firm. We consider simply the equilibrium contract in a given period, conditional upon the current level of capacity. The full dynamic equilibrium of the model is assumed to exist and to yield a value function $q(\cdot)$ that is strictly decreasing, as in the previous model (Proposition 1). We take as an equilibrium concept the usual notion of perfect competition in contract offers: firms take as given the market level of utility offered consumers, believing that they could sell an arbitrary num-

[24] Another explanation of coverage limits on insurance contracts is risk aversion of insurers; but in the standard theory of insurance markets (Section 2) insurers should be risk-neutral because risks are diversified by shareholders.

ber of any contract yielding this utility, and also take as given the (random) next period price of equity, q_{t+1}. The set of contracts from which firms choose an offer consists of pairs $[P, I(L)]$, where P is the premium paid *ex ante* and $I(L)$ is the coverage paid when the loss L is realized.

This framework yields as an equilibrium contract the one maximizing expected utility subject to two constraints: (1) *individual (firm) rationality* requires that the contract yield an expected next-period value for a competitive firm that is no less than that if the firm offered no contracts at all, simply investing in the riskless asset, and (2) the limited-liability constraint. The first constraint reduces to a simple expression. If a single competitive firm, with net worth x, simply invests in the riskless asset then its next-period value is $x(1 + r - c)q_{t+1}$. If instead it offers n contracts then its net worth next period is $\{x + n[P - pI(L)]\}(1 + r - c)$, which is multiplied by q_{t+1} to yield the next-period value.[25] The first constraint is therefore

$$Eq_{t+1}\{x + n[P - pI(L)]\}(1 + r - c) \geq Eq_{t+1}x(1 + r - c), \quad (16)$$

which (parallel to the case of uncertain p analyzed in the model of Section 2) reduces to $P \geq pEI(L) + p\,\text{cov}(q_{t+1}, I(L))/Eq_{t+1}$. This constraint is independent of n. In sum, the equilibrium contract $[P^*, I^*(L)]$ solves

$$\max_{P, I(L)} (1 - p)U(W - P) + p \int U[W - P - L + I(L)]\, dF(L) \quad (17)$$

subject to firm rationality,

$$P \geq p \cdot EI(L) + p\,\frac{\text{cov}[q_{t+1}, I(L)]}{Eq_{t+1}}, \quad (18)$$

and limited liability,

$$(\forall L) \qquad N \cdot [pI(L) - P] - X \leq 0. \quad (19)$$

In addition to this maximization problem, the equilibrium conditions include a transition equation and a recursive valuation equation.

To put this problem in perspective, changing to a perfect capital market framework would involve dropping Eq. (19) and the last term of Eq. (18). The resulting perfect capital market solution would involve $I(L) = L$, i.e., full insurance. Proposition 6 characterizes the optimal coverage here:

[25] This assumes that each contract is small, e.g., that consumers are a continuum, so that the relevant law of large numbers holds.

PROPOSITION 6. *Suppose that an equilibrium exists, with q^* strictly decreasing. Then, for some $\hat{L} > 0$, the equilibrium contract $[P^*, I^*(L)]$ is characterized by*

(*a*) $I(L) < L$ *for* $L > \hat{L}$.
(*b*) $I(L) > L$ *for* $L < \hat{L}$.

The limits on coverage of large losses result from two factors. The first is that limited liability may constrain the maximum coverage that can be offered at a fixed premium, and a consumer will not choose a contract with a premium sufficiently large to eliminate this constraint unless risk aversion is infinite. But even if the limited-liability constraint is not binding, i.e., even in a market with excess capacity, coverage limits are predicted. This is because the monotonicity of q and the correlation of losses across consumers imply that the opportunity cost to an insurer of coverage is increasing in the amount of coverage. A large claim is paid in the event that next period's capacity is small, and—since the value function $q(\cdot)$ is decreasing in capacity—in this event the marginal opportunity cost of paying the claim is high.

The prediction (b) of the proposition is surprising. The optimal sharing rule does not specify simply full coverage of losses up to some limit; instead, $I^*(L)$ *exceeds* L for low values of L (and is less than L for large values of L). As a result, the best outcome for the consumer under the contract is to realize a small loss, rather than no loss at all. This "negative deductible" is not actually observed; a more realistic interpretation is that we should see deductibles less often on contracts with coverage limits than on contracts without such limits.[26] Within this framework, a reasonable additional assumption is that $I(L)$ cannot exceed L at any L because of a moral hazard problem—that the insured could at zero effort cause a loss. With this assumption, the optimal contract covers all losses up to some ceiling that is less than the maximum possible loss. This is a form of contract that is often observed. In addition to predicting this contract form, this extension of the limited-liability model can explain (Proposition 7, below) the observation that premiums (per dollar of maximum cover-

[26] The prediction that small losses are overinsured follows from a missing market. The consumer can transfer wealth between the contingent *events* of "accident" of some size and "no accident" at a fair premium (since the events of losses are independent across consumers). Because of this the optimal contract will equate the marginal utility of wealth given no accident to its *average* value in the accident event. The marginal utility of wealth is, at the optimum, high in the subevent of a high-loss accident (because of limited coverage). The optimum condition therefore requires a very low marginal utility of wealth in the subevent of a low loss at the optimum, and this low marginal utility is achieved by overinsurance. The missing market giving rise to this effect is for the transfer of contingent wealth across the subevents of low and high accident cost, between policyholders and stockholders of insurance firms. This market is precluded by the limited-liability constraint.

age) are increasing in the amount of maximum coverage. This observation is inconsistent with standard insurance theory with risk-neutral insurers.

Consider a competitive insurance market as above, but in which contracts are restricted to those which indemnify all losses up to some ceiling C. Consumers have the same random losses, but may have heterogeneous preferences. Define the competitive equilibrium premium schedule $P(C)$ as the premium that insurers offer, in equilibrium, for different amounts of coverage; the offer of any contract on this schedule leaves the value of the insurer's equity unchanged, given state-contingent future prices of equity.[27]

PROPOSITION 7. *Suppose that contracts in the limited-liability model with varying losses are constrained to satisfy $I(L) \leq L$, and suppose that consumers have heterogeneous preferences. Then each equilibrium contract offered covers all losses up to some ceiling. Furthermore,*

(*a*) *A "fair" premium schedule $P(C)$ would satisfy $\partial(P/C)/\partial C < 0$.*

(*b*) *In the limited-liability model with uncertain losses, the equilibrium price schedule offered to consumers satisfies $\partial(P/C)/\partial C > 0$ when the limited-liability constraint is binding.*

Nonlinear pricing contracts are not a prediction of the model in Section 2, in which there is uncertainty only in the probability or frequency of losses. Within our general limited-liability approach, therefore, uncertainty in the size of losses is *necessary* to explain the market performance during the 1985 and 1986 crisis.

Uncertainty in losses, however, is not by itself *sufficient* to explain the main features of the market dynamics, in particular the disappearance of transactions in some lines. If this type of uncertainty alone is combined with the other extension of the previous section, the multiple-lines assumption, then there are always gains to trade in every line:

PROPOSITION 8. *With multiple lines, and dependence in the size of losses, but not in the events of losses, then there are transactions in every line, whatever the current capacity.*

The proof of this proposition is simple. There are always gains to trade in insuring at least the minimum possible loss because an insurer offering this loss has a perfectly predictable claim with the large number of insureds. The importance of the proposition is in the basic goal of tracing the aspects of market equilibrium back to fundamental market conditions: Within the limited-liability framework, dependence in the events of losses

[27] Given homogeneous consumers, only a single contract on this schedule is observed in equilibrium; with a variety of preferences among consumers, many contracts would be observed.

is necessary to explain the "availability" aspect of insurance crises, just as dependence in the size of losses is necessary to explain the problem of "adequacy" or tight coverage limits in equilibrium.

To link the performance of liability insurance markets to developments in underlying tort law, this means that uncertainty in the *damage awards* by courts leads to tight limits on insurance coverage. Uncertainty in liability *standards* by courts leads to the absence of transactions or gains to trade in some lines during tight markets.

5. EVIDENCE

5.1. *The Main Testable Implication*

The limited-liability or constrained-capacity model has a variety of implications that can be tested with time series-data. I test the most important implication with aggregate property-liability insurance data and then discuss evidence on other implications. Grøn (1989, 1994) offers an independent development and tests on several separate lines of the market.

The most basic implication is a rejection of the present value hypothesis, or model of insurance pricing. This hypothesis implies that premiums are sufficient statistics, within the set of information available at the date that policies are issued, for the present value of claims. The error in premiums as predictors of subsequent cash flows must be independent of any information available to the market. In other words, any change in premiums must be attributable entirely to a change in the distribution of losses or a change in interest rates. The implication of the capacity-constraint hypothesis, in contrast, is that the prediction error in premiums is negatively related to the stock of surplus at the time policies are issued.

I adopt the following assumptions on functional forms as a general empirical framework within which the two hypotheses appear as particular parametric restrictions. Let P_t represent the premiums on policies issued in year t, and $L_{t,\tau}$ represent the claims (and expenses) realized for year $\tau \geq t$ on these policies. Let PV be the present value operator on the sequences of losses $[L_{t,t}, \ldots, L_{t,t+n}]$, let θ_t be information available to the market in year t, let X_t/D_t ($\in \theta_t$) represent net worth relative to the level of demand and let ε_t be a random error term.

$$P_t = \mathrm{PV}(L_{t,t}, \ldots, L_{t,t+n})(X_t/D_t)^{-a}e^{-\varepsilon_t}. \tag{20}$$

Within this model, we are testing the present value hypothesis as a null hypothesis against our alternative. Specifically, for the present value hypothesis, $a = 0$, and for the capacity-constraint hypothesis, $a > 0$.

The present value hypothesis implies $a = 0$ because $X_t/D_t \in \theta_t$ and $P_t = E[\text{PV}(L_{t,t}, \ldots, L_{t,t+n})|\theta_t)$. Let R_t be the ratio $\text{PV}(L_{t,t}, \ldots, L_{t,t+n})/P_t$, which I term the realized *economic loss ratio* (ELR). It is convenient to rewrite (20) as

$$R_t = (X_t/D_t)^a e^{\varepsilon_t}. \qquad (21)$$

5.2. *Data and Test Methodology*

An ideal data set on which to test our model would include the time series of premiums on policies written at each date, P_t, and the flow of subsequent, realized claims, $L_{t\tau}$ for $\tau \geq t$. This would allow calculation of the realized present value of claims, using as discount rates the term structure of interest rates at t. The ideal data would also include the surplus or capacity of the world insurance market, including primary and reinsurers.

The actual data are annual for the aggregate U.S. property-liability market over the period 1948–1988. They depart from the ideal data in two ways. First, with regard to the dependent variable in Eq. (21), we do not have cash-flow data. The cash-flow data on claims paid, which are reported to state insurance commissions as "accident year losses, developed," are available for all major lines only after 1980. As an alternative, I use proxies for: (a) the proportion of claims arising, from policies sold in year t, in each year τ, $\tau \geq t$; and (b) the total nondiscounted claims from policies sold in each year t, as a ratio of premiums on those policies. Given proxies for (a) and (b), I calculate the present value of claims for policies sold in year t as a ratio of premiums, i.e., the economic loss ratio which is the dependent variable in Eq. (21). I use five-year government bond rates to discount, assuming that the underwriting risk is nonsystematic.[28] Appendix 2 describes in more detail the regression variables. The second departure from the ideal data set is that the independent variable does not include a measure of capacity of the world reinsurance market, but is instead constructed from the real stock insurers' surplus in the U.S. property-liability insurance market. This is depicted in Fig. 5.

The independent variable in the test is the *cyclical component* of the real surplus, calculated as the ratio of current surplus to its five-year historical average.[29] That is, denoting surplus by X_t, the independent variable is defined by $C_t \equiv X_t/[\frac{1}{5}\sum_{i=1}^{5} X_{t-i}]$. In this formulation, the five-

[28] Cummins and Harrington (1988) find that underwriting betas are very unstable, and may have been negative in the late 1970s. The results of the current paper were not substantially affected by the inclusion of a risk premium.

[29] "Current" means at the beginning of a calendar year, i.e., on the balance sheet on January 1, and is recorded in Best as the surplus of Dec. 31 of the previous year.

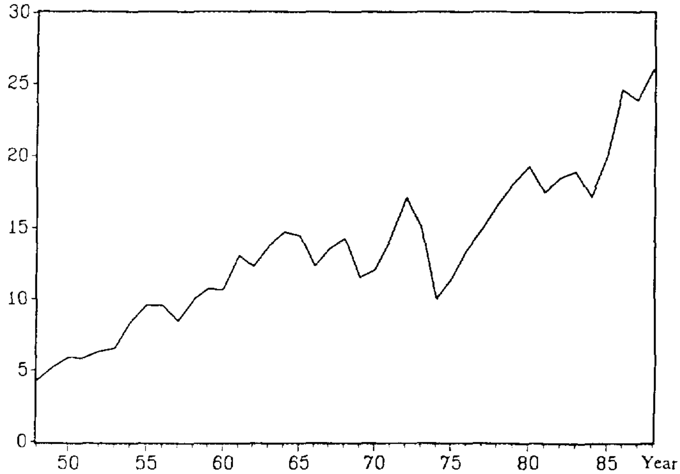

FIG. 5. Surplus of U.S. property-liability stock insurers (billions of 1967 dollars). From "Best's Averages and Aggregates."

year average is essentially a proxy for demand; it is the ratio of surplus to demand that is the ideal independent variable. The historical average is used rather than a centered, moving average to avoid spurious correlation from the inclusion of future surplus terms on the right-hand side of regression. The dependent variable, the economic loss ratio, obviously affects future surpluses.

The dependent variable and the independent variable are obviously subject to measurement error. This is not an issue for the independent variable since the null hypothesis implies that the error in premiums as predictors of subsequent claims should be independent of *any* available information (including accounting measures of capacity) at the time of issue of the policies; the test of the null hypothesis remains valid. But the test must take account of a structural change in the estimated relationship that is attributed to the measurement error in industry capacity, as discussed below. To the extent that the measurement error in the dependent variable is correlated with the accounting measure of capacity, however, the test is biased. The most likely source of such correlation is through the common factor of interest rates,[30] as net worth is measured as a book value (and is therefore surely less responsive to interest rates than market values) and the assumed maturity structure and size of losses used in calculating the dependent variable are approximations. I include current and lagged interest rates to eliminate this source of partial correlation

[30] I am grateful to a referee for pointing this out.

between the dependent and independent variables. In addition, the independent variable, C, enters the equation only with lags, so that spurious correlation with the economic loss ratio is not an issue.

The role of interest rates in the regression requires some elaboration, since they are ubiquitous in empirical tests of the insurance cycle. In these tests, the usual dependent variable is the (nondiscounted) loss ratio or underwriting profit. Assets of insurance companies are largely nominal assets such as bonds as opposed to real assets; the liabilities are real with nominal limits on coverage. Ignore, for simplicity, the real components of the balance sheet and assume that assets and liabilities are entirely nominal. Exogenous nominal interest rate changes will then have two effects on insurance loss ratios: an increase in interest rates, anticipated or not, will lower the loss ratio—under both our null hypothesis and our alternative (capacity-constraint) hypothesis. In addition, an unanticipated increase in interest rates will lead to drop in net worth of insurers to the extent that assets are of longer maturity than liabilities, and therefore—only under the alternative hypothesis—a drop in loss ratios. In our specification, the first effect is incorporated or controlled for with the use of discounted losses in the loss ratio; the second effect is controlled for since it works through changes in capacity, which is already in the regression. In short, interest rate changes will have explanatory power in the regression only to the extent of measurement error, and a zero coefficient on interest rates can be regarded as a specification or measurement-accuracy test.[31]

5.3. *Empirical Results*

The prediction that the economic loss ratio is positively correlated with the cyclical component of surplus is supported by the evidence (based only on domestic capacity) up to the 1980s. The evidence is provided by the regression estimates of Table 1. The first regression shows that the cyclical component of surplus, lagged one year, is statistically significant, with a T statistic of 5.98; the null hypothesis can be rejected at the 1% significance level against the alternative of the capacity-constraint hypothesis.

The second regression shows that capacity enters significantly with both one- and two-year lags; again the implication of the present value hypothesis that these two coefficients are zero is rejected at the 1% level.

[31] When one considers real as well as nominal assets and liabilities, one must distinguish between real interest rate shocks and price level shocks; for example, unanticipated inflation will have a negative impact on net worth to the extent that the nominal component of assets exceeds the nominal component of liabilities. Grøn (1994) incorporates unanticipated interest rate changes, inflation, and unanticipated losses as instruments for capacity or net worth changes in the industry.

TABLE I
ECONOMIC LOSS RATIO REGRESSION ESTIMATES

	Period (regression)				
	1948–1980 (1)	1949–1980 (2)	1948–1988 (3)	1949–1988 (4)	1949–1988 (5)
Constant	−0.144 (0.014)	−0.134 (0.014)	−0.144 (0.013)	−0.139 (0.013)	−0.139 (0.015)
$\log C_{t-1}$	0.216 (0.036)	0.094 (0.043)	0.157 (0.041)	0.065 (0.051)	0.097 (0.044)
$\log C_{t-2}$		0.128 (0.042)		0.114 (0.049)	0.130 (0.043)
r_t	−0.0035 (0.0075)	−0.0036 (0.0075)	−0.0079 (0.0045)	−0.0067 (0.0044)	0.0009 (0.0058)
r_{t-1}	0.0109 (0.0079)	0.0096 (0.0077)	0.0172 (0.0045)	0.0153 (0.0044)	0.0076 (0.0059)
$d \log C_{t-1}$					−0.313 (0.142)
$d \log C_{t-2}$					−0.288 (0.150)
dr_t					0.0036 (0.0046)
dr_{t-1}					0.0064 (0.0044)
MA(1)	0.956 (0.191)	0.885 (0.192)	0.769 (0.167)	0.789 (0.169)	0.841 (0.184)
adj R^2	0.639	0.653	0.654	0.639	0.752
SER	0.032	0.031	0.041	0.039	0.032
SSR	0.029	0.025	0.059	0.05078	0.0308

Note. Dependent variable, log ELR; SER, standard error of the regression; SSR, sum of squared residuals; MA(1), coefficient estimate for moving average.

In both regressions, interest rates have only minor additional explanatory power, with a moderate change in the interest rate having a negligible impact on the economic loss ratio. This is consistent with there being little measurement error in our variables, as explained above. In all the regressions, a moving average process fit the error structure better than an autoregressive process; there is no a priori reason to presume the latter.[32]

[32] Under the null hypothesis of informational efficiency, the errors should in fact exhibit zero serial correlation. I don't focus on this test of the theory, since it is too sensitive to measurement error.

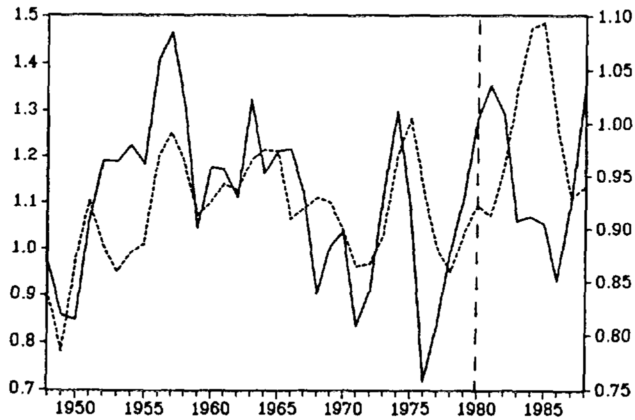

FIG. 6. Economic loss ration (ELR) and cyclical component (C) of surplus, 1948–1988. ---, ELR; —, C.

The economic loss ratio and the (optimally weighted average of the lagged, cyclical component of) capacity are depicted in Fig. 6. This figure is useful for tracing the history of the property-liability insurance market in the postwar period. Referring to the solid line, between the mid-1950s and 1984 there were four major episodes of tight capacity. According to Stewart's (1984, pp. 304–305) history of the postwar market, the tight markets in 1957 and 1969 were in automobile insurance lines and followed exceptionally bad underwriting results, each time set off by an episode of high inflation, and the crunch in 1965 was due to the loss of reinsurance capacity following the property damage with a string of major hurricanes. The losses of the mid-1970s were capital losses on long-term bonds as interest rates rose suddenly. As the figure shows, these tight capacity episodes are all associated with high premiums relative to costs (low economic loss ratios), and the high loss ratios (or low premiums) of the soft markets in the mid-1950s and the early 1970s are both associated with a large cyclical component of surplus.

While the main implication of the limited-liability model is supported by the pre-1980s evidence, the simple model cannot capture fully the premium dynamics. Because of the simplifying assumption of one-period contracts, with no overlapping contracts, the theoretical model yields the prediction that premiums (and the economic loss ratio, given the stationary claims distribution) follow a first-order Markov process. The data reveal a *second-order* autoregressive process for the economic loss ratio, as previous authors (e.g., Cummins and Outreville, 1987) have found for the accounting loss ratio. The estimated process is

$$R_t = 0.58 + 0.83R_{t-1} - 0.46R_{t-2}. \tag{22}$$

The standard errors of the three coefficients in Eq. (22) are 0.11, 0.14, and 0.14, respectively; the R^2 is 0.51. A second-order difference equation such as (22) yields a cycle if the coefficients a_1 and a_2 satisfy $a_1^2 + 4a_2 < 0$, and the period of the cycle is given by $2\pi/\cos^{-1}(a_1/2\sqrt{-a_2})$. In our case, the cycle is dampened (since $\sqrt{-a_2} < 1$) with a period of 6.8 years.[33] Furthermore, the estimates from regression 2 of Table 1 show that capacity enters with a one- and a two-year lag, while the theoretical model predicts that the current level of capacity is the single state variable. A one-year lag can be explained on the basis of the form of the data, but a two-year lag cannot.

An extension of the model to incorporate overlapping policy maturity periods would yield a higher-order process for premiums, as well as a higher-dimensional state space. In such an extension, the entire maturity structure of liabilities of the market would compose the state variables. The "right" model of insurance premium dynamics would incorporate as well a stochastic drift in the distribution of common factors (similar to the representation of technology in many real business cycle models). In the actual market, insurers are stuck with the liabilities of previous policies, possibly from the distant past, when tort law changes.[34]

5.4. *The 1984–1986 Experience*

The economic loss ratio and the cyclical component of capacity of domestic insurers were *negatively* related to the 1980s. The correlation between these variables changed from 0.53 before 1980 to −0.64 afterward. Premiums fell relative to discounted claims in the early 1980s, even as capacity of primary insurers fell. Then, over the 1984–1986 period, premiums rose relative to claims and capital entered the industry. Table II shows that the capital raised by the industry changed from net exit in the early 1980s to entry in 1984.

The negative relation between the economic loss ratio and measured capacity during the early 1980s may reflect simply a series of unanticipated losses over that period; the measured losses are *ex post* and nothing in the theory contradicts a run of bad luck. Statistically, however, the

[33] See Cummins and Outreville (1987) for similar findings on accounting-loss ratios. There is some evidence of structural change in about 1980 (as explained below). For this reason, I estimated the process over the period 1948–1980.

[34] For example, in the asbestosis cases, damages were often manifest only 20 or 30 years after the accident, by which time tort law had evolved to the point where it assigned liability to third parties for even unknowable risks, and to any insurer covering the party between the accident and the suits (*Beshada v. Johns-Manville Prod. Corp.*, 90 N.J. 191, 447 A.2d 539 (1982)). In other litigation, involving a drug (DES) taken by pregnant women, accident damages were manifest more than 15 years after ingestion of the drug (see Trebilcock, 1987).

TABLE II
NET CAPITAL AND SURPLUS PAID IN, U.S. PROPERTY-LIABILITY INSURANCE, 1980–1986

Year	Capital and surplus paid in, net of dividends to shareholders (billions)
1980	−1.495
1981	−1.775
1982	−1.248
1983	−1.945
1984	0.322
1985	5.561
1986	1.300

hypothesis of zero structural change in 1980 can be rejected at the 1% significance level.[35]

The structural change is consistent with a change in the *source of shocks* in the 1980s, in combination with an omitted variable in the regressions. Measured capacity is capacity of primary U.S. insurers only; the capacity of the reinsurance market is omitted. The capacity of the reinsurance market was large in the early 1980s, to some extent because a change in U.S. tax law artificially encouraged entry into the reinsurance market.[36] 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 consequence was that the high claims in the early 1980s, combined with artificially low premiums, led to insolvency of many of the new entrants. Between 1984 and mid-1987, some 90 reinsurance companies left the market. The rate of insolvencies among primary insurers tripled just prior to 1984 crisis, as it had in 1975 (see Winter, 1988). A driving force behind the 1980s soft market and subsequent crisis appears to be these changes in the capacity in the reinsurance market.

The limited-liability model predicts that eventually a tight market will soften. This happened with the mid-1980s crisis. Premiums fell in many

[35] This hypothesis is tested using the standard method, by interacting the variables in the regression with a dummy variable, d, taking on the value of 1 after 1980 and 0 up to 1980 in regression (5) of Table I and testing that the coefficient on the interaction term is zero.

[36] The entry was by *captives*, which are upstream insurance corporations established by a company or group of companies for the purpose of providing insurance. Captives provided tax advantages to parent companies over direct self-insurance. The tax-favored status of captives as *bona fide* insurers was tenuous in the early 1980s, however, following changes in the common law interpretation of the tax code. Captives entered the reinsurance market as well as the risky excess-and-surplus line in the United States on the theory that this would establish them as *bona fide* insurers. See Winter (1988).

lines in 1987 and 1988. In the most uncertain lines (hence the lines most sensitive to changes in capacity), which are surplus and excess coverage lines, some premiums were cut by 40% in 1987 and by 20% more in 1988 (see *Business Insurance*, p. 25, July 4, 1988). Conventional or "admitted" insurers cut into the surplus and excess coverage market by increasing limits on basic coverage in some cases by 500%. In other words, the market price not only fell but became flatter as a function of coverage, as predicted in Section 4.

5.5. *Additional Testable Implications*

The limited-liability model offers additional predictions. The prediction that capital should enter mainly during tight market periods is consistent with the evidence, as indicated for recent data in Table II. This distinguishes the model from the hypothesis that insurers exited the market during the most recent crisis simply because of rapidly rising costs of coverage.

The limited-liability model predicts that the market-to-book ratio for insurers should be a declining function of capacity. In addition, under the assumption of low barriers to entry of equity, and substantial barriers to exit of capital ($k_1 = 0$ and $k_2 > 0$) in the formal model—for example, that the trapped-equity model of dividend taxation is the driving force—the equilibrium value of market-to-book ratio, q, must be less than one. Hence the stock market reaction to a sudden, unanticipated loss is a reduction in market value of less than the dollars paid out in claims. Grøn (1989) tests this implication using as experiments the San Francisco earthquake of October 7, 1989, and Hurrican Hugo two weeks earlier and finds convincing evidence of the capacity-constraint hypothesis.

Finally, the extension to multiple lines shows that those lines whose claims are most sensitive to aggregate common factors should also exhibit the highest sensitivity in pricing to changes in aggregate capacity. Grøn, again, finds strong support for the capacity-constraint hypothesis using individual line data, but the cross-sectional implication has yet to be tested. Future models of insurance pricing, toward which the current literature has taken some steps, will explain comovements in a large set of market variables including premiums, expected claims, coverage, capital entry and exit of capital, and the stock market value of insurers.

6. CONCLUSION

Was the sudden increase in premiums in 1985 an actuarially justified response to an explosion in tort awards or a collective decision by the industry to justify higher premiums by exaggerating the torts explosion? This paper suggests that both sides of this debate were wrong. The increase in premiums was greater than "justified" by expected claims on the one hand, and on the other, this is entirely consistent with a competitive insurance market.

The conventional economic theory of competitive insurance markets assumes that premiums predict efficiently future claims. This perfect capital market theory is implicit in almost all of the policy literature on the recent insurance crisis, but does not explain the stylized facts of the crisis. It is also, in this paper, rejected in a test against the limited-liability model using postwar data.

Using the limited-liability mode, we traced the features of the recent insurance market performance to basic market conditions combined with tight capacity: the disappearance of gains to trade in lines such as municipality and day-care liability resulted from dependence among the events of losses. Nonlinearity of premiums was due to dependence among the sizes of losses, conditional upon the events of losses. A goal of the law and economics literature in this area (e.g., Priest, 1987; Trebilcock, 1987) has been to link the performance of the liability insurance market to underlying shifts in tort law, especially the expansion of liability standards and awards by courts. According to the limited-liability model the uncertainty or volatility of tort law is the key variable in this link. The withdrawal of insurers from some lines in tight markets is linked to uncertainty in liability standards, and nonlinearity of pricing is linked to uncertainty in the size of tort awards.

APPENDIX 1: PROOFS OF PROPOSITIONS

Proof of Proposition 5

Let 1, 2 be two lines with $a_1 > a_2$. Then

$$\begin{aligned} a_2 - a_1 < 0 \Rightarrow (a_2 - a_1)\bar{p} < 0 \Rightarrow 0 > a_2\bar{p} + a_1a_2(p_{\mathrm{H}} - \bar{p}) - a_1\bar{p} \\ - a_1a_2(p_{\mathrm{H}} - \bar{p}) = a_2[\bar{p} + a_1(p_{\mathrm{H}} - \bar{p})] - a_1[\bar{p} + a_2(p_{\mathrm{H}} - \bar{p})] \\ = a_2p_{\mathrm{H}1} - a_1p_{\mathrm{H}2}. \end{aligned} \quad (23)$$

Now, denoting $\bar{q}_{t+1} \equiv E\tilde{q}_{t+1}$, we have from (15) of the text

$$\begin{aligned} (p_{\mathrm{H}i} - P_i) &= p_{\mathrm{H}i} - \frac{1}{\lambda + \bar{q}_{t+1}}[\lambda p_{\mathrm{H}i} + a_i\mathrm{cov}(q_{t+1}, p) + \bar{q}_{t+1}\bar{p}] \\ &= \frac{\bar{q}_{t+1}p_{\mathrm{H}i} - a_i\mathrm{cov}(q_{t+1}, p) - \bar{q}_{t+1}\bar{p}}{\lambda + \bar{q}_{t+1}} \\ &= \frac{\bar{q}_{t+1}[\bar{p} + a_i(p_{\mathrm{H}} - \bar{p})] - a_i\mathrm{cov}(q_{t+1}, p) - \bar{q}_{t+1}\bar{p}}{\lambda + \bar{q}_{t+1}} \\ &= a_i\left[\frac{\bar{q}_{t+1}(p_{\mathrm{H}} - \bar{p}) - [E(q_{t+1}p) - \bar{q}_{t+1}\bar{p}]}{\lambda + \bar{q}_{t+1}}\right] \\ &= a_i\left[\frac{E[q_{t+1}(p_{\mathrm{H}} - p)]}{\lambda + \bar{q}_{t+1}}\right]. \end{aligned} \quad (24)$$

Hence, across lines $(p_{Hi} - P_i)$ is proportional to a_i, which implies

$$\frac{(p_{H1} - P_1)}{(p_{H2} - P_2)} = \frac{a_1}{a_2} \Rightarrow a_2 P_1 - a_1 P_2 = a_2 p_{H1} - a_1 p_{H2}. \tag{25}$$

Equations (23) and (25) imply $a_2 P_1 - a_1 P_2 < 0$, which implies

$$\frac{a_2}{a_1}\frac{P_1}{P_2} < 1. \tag{26}$$

Since $P_1 > P_2$ by Proposition 4, and the demand elasticity (denoted ε_i at each P_i) is increasing in P, we have

$$\varepsilon_1/\varepsilon_2 > 1. \tag{27}$$

From (25), with a change dX in total capacity, we have $dP_2 = (a_2/a_1)dP_1$, which implies

$$\frac{d \ln P_2}{d \ln P_1} = \frac{a_2}{a_1}\frac{P_1}{P_2}. \tag{28}$$

Now, (26), (27), and (28) imply

$$\varepsilon_1 \frac{d \ln P_1}{d \ln X} - \varepsilon_2 \frac{d \ln P_2}{d \ln X} > 0. \tag{29}$$

Next, (25) and the limited liability constraint (12) imply

$$\frac{X_1/Q(P_1)}{X_2/Q(P_2)} = \frac{a_1}{a_2} \Rightarrow \frac{X_1}{Q(P_1)} = \frac{a_1}{a_2}\frac{X_2}{Q(P_2)}$$
$$\Rightarrow \frac{d \ln X_1}{d \ln X} - \frac{d \ln Q(P_1)}{d \ln X} = \frac{d \ln X_2}{d \ln X} - \frac{d \ln Q(P_2)}{d \ln X}.$$

This implies

$$\begin{aligned}\frac{d \ln X_1}{d \ln X} - \frac{d \ln X_2}{d \ln X} &= \frac{d \ln Q(P_1)}{d \ln X} - \frac{d \ln Q(P_2)}{d \ln X} \\ &= \frac{d \ln Q(P_1)}{d \ln P_1}\frac{d \ln P_1}{d \ln X} - \frac{d \ln Q(P_2)}{d \ln P_2}\frac{d \ln P_2}{d \ln X} \\ &= \varepsilon_1 \frac{d \ln P_1}{d \ln X} - \varepsilon_2 \frac{d \ln P_2}{d \ln X}.\end{aligned} \tag{30}$$

Equations (29) and (30) imply

$$\frac{d \ln X_1}{d \ln X} > \frac{d \ln X_2}{d \ln X},$$

proving Proposition 5.

Proof of Proposition 6

The competitive firm takes the future price q under each realization of L as given, not recognizing the impact on q of its contribution to X. With abuse of notation, I denote this anticipated price by $q[L]$. Let $f(\cdot)$ be the density F'.

In solving the maximization problem (17), it is convenient to reexpress the constraint (18) as

$$p \int q[L]I(L)f(L)dL - PEq \leq 0. \tag{31}$$

To save space, I consider here the case of a nonbinding limited-liability constraint, showing that even in this case the optimal contract does not offer full coverage. The full problem yields the same results. The problem of maximizing (17) over the space of sharing rules subject to (31) is a problem of Lagrange, and the solution is given by first-order conditions on the point-wise choice of $I(L)$. Letting the shadow price on the constraint be λ, the necessary conditions for the problem are given by

$$-(1 - p)U'(W - P) - p \int U'[W - P - L + I(L)]f(L)dL - \lambda Eq = 0, \tag{32}$$

and, for every L,

$$pU'[W - P - L + I(L)]f(L) + \lambda pq[L]f(L) = 0,$$

which reduces to

$$U'[W - P - L + I(L)] - \lambda q[L] = 0. \tag{33}$$

Because q^* is strictly decreasing in X and the next period's capacity is decreasing in L, q^* is decreasing in L. From the concavity of U, U' is decreasing in wealth. Hence (33) implies that $I(L) - L$ is increasing in L, or that the realized wealth, $\omega(L) \equiv W - P - L + I(L)$ is increasing in L.

Hence

$$dI^*/dL > 1. \tag{34}$$

From (32),

$$\lambda = EU'/Eq, \tag{35}$$

where EU' is the unconditional expectation of U' at the optimum. Integrating (33) with respect to the distribution F of L shows that

$$\lambda = E(U'|\, L > 0)/Eq. \tag{36}$$

Equations (35) and (36) imply that $E(U'|\, L > 0) = EU'$, whence

$$E(U'|\, L > 0) = E(U'|L = 0) \equiv U'(W - P). \tag{37}$$

Finally, the montonicity of $\omega(L)$, (37), and the monotonicity of U' imply directly that $I(L) > L$ for low L and $I(L) > L$ for high L. There is "overinsurance" of small losses and less than full coverage of large losses at the optimum.

Proof of Proposition 7

(a) The actuarially fair premium satisfies

$$P = \int_0^M Lf(L)dL + M[1 - F(M)].$$

Dividing by M and differentiating yields

$$\frac{\partial(P/M)}{\partial M} = f(M) - (1/M^2)\int_0^M Lf(L)dL - f(M) < 0.$$

(b) If the limited-liability constraint is binding on the offer of a contract to n consumers, then $X + n(P - p_H M) = 0$, which implies $P/M = p_H - X/(nM)$ and

$$\frac{\partial(P/M)}{\partial M} = \left(\frac{X}{n}\right)\left(\frac{1}{M^2}\right) > 0.$$

APPENDIX 2: CONSTRUCTION OF REGRESSION VARIABLES

This appendix describes the variables constructed for the empirical test of the model. The data are for aggregate stock property-liability insurers in the United States.

Real Surplus

Real surplus in year t is measured as the aggregate surplus or net worth on the balance sheets of U.S. stock insurers on Dec. 31 of the previous year, deflated by the CPI.

Economic Loss Ratio

This variable is constructed from accounting data on aggregate stock insurers' "losses" (claims paid) and premiums. The main input is the time series of *loss ratios*: the numerator of this ratio is the claims paid by an insurer during a year, plus the increase in a reserve representing the nondiscounted future claims to be paid from policies already written. We take this numerator as a proxy for the realized claims during the year of new policies, plus the expected nondiscounted claims to be paid on those policies in the future. (We ignore the fact that losses in any year include revisions in the reserves representing future claims from policies written in previous years.) The denominator of this ratio is net premiums earned. These differ from net premiums written in that, for example, only one-quarter of the premium paid for a policy written on Sept. 30 would be included in the current year, with the remainder of the premium allocated to earned premiums of the next year. While net premiums *written* is the variable that should be responsive to current economic conditions (such as a change in industry capacity), to ensure that the denominator and the numerator correspond more closely to the same policies, it is necessary to use net premiums earned. The second input is the "expense ratio," which gives the variable expenses other than claims (expenses such as commissions) incurred by stock insurers, as a ratio of premiums. We take as an approximation that these expenses are incurred contemporaneously with premiums. Dividing the loss ratio, L, by 1 minus the expense ratio, e, gives a proxy for nondiscounted claims as a fraction of (premiums net of expenses) $L/(1 - e)$. The loss ratios and expense ratios were obtained from "Best's Averages and Aggregates."

Let β_s represent the proportion of claims paid in each year afer an insurance policy is written. Multiplying $L/(1 - e)$ by $D \equiv \Sigma\ \beta_s/(1 + r)^s$ yields an estimate of the present discounted value of claims as a fraction of premiums net of expenses. The estimates for the vector β_s were determined as follows. For "Schedule P" lines, which involve relatively long-tailed losses, cash-flow data on the cumulative claims paid in each year, subsequent to the policy year, are available for each year starting in 1980 (e.g., Bests, 1989, p. 76). Differencing this series yields an estimate for β. For Schedule P lines, I averaged the resulting estimates of β for the three years 1980–1982 (over which there was very little variation). For "Schedule Q" lines, which do not involve long-tailed losses, the proportion of claims in each year was, in 1986, approximately 60% in the first year, 30% in the second, and 10% thereafter. This is approximated to a series (0.6, 0.3, 0.1, 0, . . .). Finally, the β vectors for Schedule P and Schedule Q lines are averaged and weighted by the 1986 premium revenue from each class (114 billion for P lines and 52 billion for Q lines). The resulting β vector is (0.435, 0.266, 0.107, 0.055, 0.034, 0.034, 0.034, 0.034). Note that the econometric results were not sensitive to variation in this estimate of the economic loss ratio, and that the inclusion of interest rates as regressors would pick up any significant measurement error resulting from miscalculation of the maturity structure of losses.

REFERENCES

ARROW, K. J. (1970). "Essays in the Theory of Risk Bearing," Markham, Chicago.

ASQUITH, P., AND MULLINS, D. W. (1986). Equity issues and offering dilution, *J. Finan. Econ.* **15,** 61–89.

AUERBACH, A. (1983). Taxation, corporate financial policy and the cost of capital, *J. Econ. Lit.* **21,** 905–940.

BAGWELL, L. S., AND SHOVEN J. (1989). Cash distributions to shareholders, *J. Econ. Perspect.* **3,** 129–141.

BERGER, L. A. (1988). A model of underwriting cycles in the property-liability insurance industry, *J. Risk Ins.* **55,** 298–306.

BERGER, L., CUMMINS, J. D., AND TENNYSON, S. (1992). Reinsurance and the liability insurance crisis, *J. Risk Uncertainty* **5,** 253–272.

A. M. BEST CO. (annual). "Best's Aggregates and Averages," NJ.

BRIGGS, E., DIONNE, G., AND EECKHOULT, L. (1989). More on insurance as a Giffen good, *J. Risk Uncertainty* **2,** 420–425.

CUMMINS, J. D., AND HARRINGTON, S. (1988). The relationship between risk and return: Evidence for property-liability insurance stocks, *J. Risk Ins.* **55,** 15–31.

CUMMINS, J. D., AND OUTREVILLE, J. (1987). An international analysis of underwriting cycles in property-liability insurance, *J. Risk Ins.* **54,** 246–62.

DANZON, P., AND HARRINGTON, S. (1992). The demand for and supply of liability insurance, *in* "Contributions to Insurance Economics" (G. Dionne, Ed.), pp. 25–60. Kluwer Academic, Boston.

DOHERTY, N. A., AND BANG, H. B. (1988). Interest rates and insurance price cycles, *J. Bank. Finan.* **12,** 199–214.

FAIRLEY, W. B. (1979). Investment income and profit margins in property-liability insurance: Theory and empirical results, *Bell J. Econ.* **10,** 192–210.

GREENWALD, B., AND STIGLITZ, J. E. (1993). Financial market imperfections and business cycles, *Quart. J. Econ.* **108,** 77–114.

GRØN, A. (1989). Property-casualty insurance cycles, capacity constraints, and empirical results, Ph.D. dissertation, Department of Economics, MIT.

GRØN, A. (1994). Capacity constraints and cycles in property-casualty insurance markets, *Rand J. Econ.* **25,** 110–127.

HILL, R. (1979) Profit regulation in property-liability insurance, *Bell J. Econ.* **10,** 172–191.

HOY, M., AND ROBSON, A. (1981). Insurance as a Giffen good, *Econ. Lett.* **8,** 47–51.

HUBERMAN, G., MAYERS, D., AND SMITH, C., JR., (1983). Optimal insurance policy indemnity schedules, *Bell J. Econ.* **9,** 415–426.

KRAUS, A., AND ROSS, S. A. (1982), The determinants of fair profits for the property-liability insurance firm, *J. Finance* **37,** 1015–1028.

MYERS, S., AND MAJLUF, N. (1984). Corporate financing and investment decisions when firms have information that investors do not have, *J. Finan. Econ.* **13,** 187–221.

NYE, B., AND HOFFLANDER, A. (1987). Economics of oligopoly: Medical malpractice insurance as a classic illustration, *J. Risk Ins.* **54,** 502–519.

PRIEST, G. L. (1987). Modern tort law and the current insurability crisis, *Yale Law J.* **96,** 1521–1590.

ROMANO, R. (1989). What went wrong with directors' and officers liability insurance?, *Delaware J. Corp. Law* **14,** 1–33.

ROTHSCHILD, M., AND STIGLITZ, J. E. (1976). Equilibrium in competitive insurance markets, *Quart. J. Econ.* **80,** 629–649.

SMITH, B. D., AND STUTZER, M. J. (1990). Adverse selection, aggregate uncertainty and the role for mutual insurance companies, *J. Bus.* **63,** 493–510.

STEWART, B. (1984). Profit cycles in property-liability insurance, *in* "Issues in Insurance", Vol. I. American Institute for Property and Liability Underwriters.

STIGLITZ, J. E. (1982). Information and capital markets, *in* "Financial Economics: Essays in Honor of Paul Cootner" (W. F. Sharpe and C. Cootner, Eds.), pp. 118–158. Prentice-Hall, Englewood Cliffs, NJ.

STOKEY, N., AND LUCAS, R., WITH PRESCOTT, E. (1989). "Recursive Methods in Economic Dynamics," Harvard Univ. Press, Cambridge, MA.

TORT POLICY WORKING GROUP (1986). "Report on the Causes, Extent and Policy Implications of the Current Crisis in Insurance Availability and Affordability" U.S. Department of Justice.

TORT POLICY WORKING GROUP (1987). "An Update on the Liability Crisis," U.S. Department of Justice.

TREBILCOCK, M. J. (1987). The social insurance-deterrence dilemma of modern North American tort law, *San Diego Law Rev.* **24,** 929–1002.

VENEZIAN, E. (1985). Rate-making methods and profit cycles in property-liability insurance, *J. Risk Ins.* **52,** 477–500.

WINTER, R. A. (1988). The liability crisis and the dynamics of competitive insurance markets, *Yale J. Reg.* **5,** 455–500.

WINTER, R. A. (1991a). Solvency regulation and the insurance cycle, *Econ. Inq.* **29,** 458–471.

WINTER, R. A. (1991b). The liability insurance market, *J. Econ. Perspect.* **5,** 115–136.