Capacity constraints and cycles in property-casualty insurance markets

Anne Gron*

Underwriting cycles are unexpected in a structurally competitive industry where financial capital is the major determinant of output capacity. "Arbitrage" theories explain underwriting cycles as largely an artifact of institutional lags and reporting practices; "capacity-constraint" theories view insurance markets as characterized by real frictions that cause underwriting cycles by temporarily reducing the industry's capacity to insure risks. Arbitrage theories imply no systematic relationship between capacity and underwriting margins, while capacity-constraint hypotheses predict a negative relationship. This article provides a test of the two theories by examining the empirical relationship between capacity and underwriting margins. The results, using data on four insurance lines, generally support the capacity-constraint hypothesis; unanticipated decreases in capacity cause higher profitability and prices.

1. Introduction

General liability insurance revenues increased more than 78% in 1985 and over 67% in 1986. Buyers of general liability coverage were purchasing less insurance for higher premiums. In the period 1985–1986, prices rose across most lines of insurance as they had similar, though less severe, experiences. Since 1950 there have been six such episodes of rapidly increasing prices. Insurance "crises," as these episodes are called, are one stage of a larger phenomenon known as the "property-casualty insurance cycle" or "underwriting cycle."¹

The underwriting cycle refers to a repeating series of conditions that affect property-casualty insurance markets.² The typical description of the cycle includes four stages, described in terms of industry accounting profitability. The first stage is marked by several years of low profitability. This is followed by a sudden change to rapidly increasing profitability. The second stage is often called the insurance crisis, owing to the characteristics described above. In the third stage, profitability remains high but is no longer increasing.

*University of Chicago.

I thank Paul Joskow, Nancy Rose, Jim Poterba, and Mitch Petersen for helpful conversations. This article has also benefited from the comments of seminar participants at MIT, the University of Chicago, and the Wharton School, as well as from those of two anonymous referees and the Coeditor, Timothy Bresnahan. Financial support from the Bradley Foundation and the Olin Foundation is gratefully acknowledged. Jean Lucey and Phyllis Smithers at the Insurance Library Association of Boston provided valuable research assistance.

¹ Data are from The A.M. Best Company publication Best's Aggregates and Averages.

² This article describes insurance cycles as the result of slow adjustment to random shocks, rather than a phenomenon that repeats with a particular periodicity.
Profitability gradually declines during the fourth stage as the industry returns to a period of low profitability. Anecdotal evidence suggests price and quantity movements accompany the changes in profitability. In the first stage prices are relatively low and quantity is abundant. Prices rise rapidly and quantity available drops sharply during the "crisis" stage. High prices and slowly increasing quantity follow. Falling prices and relatively abundant quantity accompany the falling profitability of the fourth stage of the cycle.

Many observers find the underwriting cycle unexpected in a structurally competitive industry where financial capital is the major determinant of output capacity. Consequently, there has been substantial interest in modeling price determination and underwriting cycles. There are two major schools of thought: "arbitrage" theory explains the presence of underwriting cycles as a phenomenon limited to accounting profitability and largely an artifact of institutional lags and reporting practices; "capacity-constraint" theory views insurance markets as characterized by real frictions and imperfections that cause underwriting cycles by temporarily reducing the industry's capacity to insure risks.

The two theories imply different relationships between the level of industry capacity and the underwriting margin. In arbitrage models the primary determinant of output capacity is financial capital (net worth) that adjusts quickly and costlessly. Arbitrage theories explain movements in underwriting margins by market features such as data collection, regulatory lags, policy renewal lags, and accounting practices. Thus arbitrage models imply no systematic relationship between capacity and underwriting margins. On the other hand, capacity-constraint models assume that net worth is subject to random shocks and real frictions that prevent it from quickly adjusting to long-run equilibrium. For example, negative net worth shocks would lead to rapid price increases. High prices erode slowly as net worth adjusts. Therefore, under the capacity-constraint hypothesis, capacity is negatively related to the underwriting margin.

This article examines the relationship between capacity and underwriting margins using industry-time-series data for the period 1952-1986. I examine whether unanticipated changes in capacity are related to changes in the underwriting margin in a framework that controls for the reporting practices identified by arbitrage theory. Four lines of insurance are considered: automobile physical damage, automobile liability, homeowners' multiple peril, and other liability. The results generally support the capacity-constraint hypothesis; unanticipated decreases in capacity cause higher future profitability and prices.

These results are important for understanding the operations of property-casualty markets. An important criterion for determining if insurance markets function well is to examine whether market prices correspond to those that would prevail in a perfectly competitive marketplace. Typically, actual insurance prices are compared to "competitive prices" predicted by arbitrage models (see, for example, Harrington (1988), Fairley (1979), and Myers and Cohn (1987)). However, prices from competitively structured insurance markets with periodic capacity shortages may not coincide with those predicted by arbitrage models. The results in this article suggest that arbitrage models are better suited for predicting long-run prices than short-run prices. Therefore, if actual prices do not correspond with prices predicted by arbitrage theory, this failure does not necessarily indicate significant noncompetitive behavior.

The article is organized as follows. Section 2 discusses capacity-constraint and arbitrage explanations of cycles in a competitively structured industry. In Section 3 I develop a specification for estimation and describe the data used in estimation. Section 4 presents the regression results and implications of the fitted equation. Finally, Section 5 summarizes the results and discusses their relevance for current issues in insurance regulation.

The models discussed in this article explain the general features of insurance cycles and not the specifics of each cycle that cause price increases to be more severe for particular insurance lines.

The underwriting margin is a "profitability" measure of the form (revenues - costs)/ revenues). Such ratios are often called underwriting profit in the insurance literature.
2. Theories of insurance cycles

- Stewart (1984) and Bloom (1987) develop two informal capacity-constraint models. They formalize the idea that "lack of capacity" causes underwriting cycles. Both models rely on shifting insurance supply curves to explain the underwriting cycle. While the level of net worth influences the position of the supply curve, both authors rely on insurer "attitudes," "expectations," and "perceptions" to explain the timing and length of the high-price phase. Winter (1988, 1989) and Gron (1989, 1990) address some of these problems. These versions of the capacity-constraint theory assume that insurers hold net worth in order to keep the probability of bankruptcy low and to maintain the ability of insurers to meet claims made by policyholders. Gron assumes this is done to comply with a regulatory requirement. Winter assumes that firms cannot become bankrupt. This generates an upward-sloping short-run supply curve whose position is determined by the level of net worth.\(^5\)

In both Winter and Gron a cost differential between internal and external sources of capital prevents financial capital from quickly adjusting when it diverges from long-run equilibrium. In particular, capital from internal sources is less costly than capital from external sources. When capacity is low, external capital will enter the industry, but firms prefer to let the accumulation of retained earnings compose part of the increase in capacity.

Under these conditions, net worth shocks generate the market conditions associated with the underwriting cycle as described below. Large adverse shocks to claims or asset values, such as unanticipated inflation, large catastrophes, unexpected increases in interest rates, or declines on the stock market, can substantially reduce industry capacity.\(^6\) Because it is costly to adjust net worth, the industry will not return immediately to long-run equilibrium. The industry supply curve shifts to the left; industry accounting profit increases, reflecting the increased opportunity cost of scarce capital. As new capital and retained earnings increase industry net worth, the supply curve shifts to the right, causing premiums and profitability to decline toward their long-run equilibrium levels.

In contrast, arbitrage theory assumes that capital adjusts quickly and costlessly. Therefore, price equals the present discounted value of costs. These conditions do not allow for persistent effects from supply shocks, and indeed, arbitrage theories of underwriting cycles explain how accounting profit cycles occur without slow adjustment to supply shocks. Venezian (1985) hypothesizes that naive loss forecasting (using a relatively simple regression methodology to predict future losses) produces underwriting cycles. Cummins and Outreville (1987) argue that regulatory and institutional lags will produce underwriting cycles in the absence of any changes in underlying supply and demand characteristics.\(^7\)

The primary difference in price determination under the two theories is summarized graphically in Figure 1. Under the capacity-constraint hypothesis the short-run premium will vary with the current level of capacity. As a result, when capacity is low the short-run supply curve shifts to the left, from \(S'\) to \(S''\), resulting in higher price and lower quantity. Net worth increases as the industry moves toward long-run equilibrium; the supply curve moves to the right—price decreases and quantity increases. \(S_{LR}\) describes the long-run supply curve. Since arbitrage theories assume that insurance supply does not depend on the short-run level of capital or that capital adjusts quickly and costlessly, \(S_{LR}\)

\(^5\) These constraints yield a relation between price and the quantity of insurance sold for a given level of net worth. Beyond some maximum quantity, the firm must receive a higher price in order to sell more insurance and not violate the regulatory or bankruptcy condition.

\(^6\) Demand shocks may also cause supply shortages, but insurance demand is generally considered relatively stable.

\(^7\) Cummins and Outreville are particularly careful in modelling accounting profitability. They recognize that reporting methods for accounting profits cause accounting profitability to reflect previous as well as current market conditions. These procedures and their effects are described in Section 3 and controlled for in the estimation.
Under arbitrage theory, price will change only if the costs (primarily noncapital) of supplying insurance change.\footnote{Other models with supply shifts include McGee (1986), where interest rates shift the supply curve, and Doherty and Kang (1988), where firms slowly adjust output. Neither model explains why the short-run supply curve slopes upwards, or what determines the timing of the cycle.}

So far in this discussion, demand is assumed to be constant. More generally, the relevant concept of capacity is one that measures the size of industry capacity relative to potential demand. I shall refer to this as relative capacity.

3. Specification and data

- The capacity-constraint theory predicts that underwriting margins will be higher when relative capacity is low. In contrast, arbitrage theory predicts no systematic relationship between the two variables. Therefore, determining whether the underwriting margin is negatively correlated with relative capacity provides a test of the two theories. In this section I develop a specification that controls for accounting and reporting features that affect the underwriting margin. I begin with a description of property-casualty insurance accounting procedures that affect the analysis. Next I describe the components of insurance profitability that determine a reduced-form equation for estimation. Finally I describe the data used in the empirical analysis of Section 4.

\footnote{Property-casualty accounting. It is well known that standard accounting measures differ from the desired economic variables. Property-casualty insurance accounting rules, known as statutory accounting practices (SAP), differ substantially from usual accounting procedures, known as generally accepted accounting practices (GAAP). The major differences that influence this analysis concern the timing of the recognition of revenues and the treatment of claims costs. These are covered here. Appendix A contains a detailed illustration of how cash flows relate to accounting measures.

Unlike standard accounting practices, SAP recognize premium revenue only as it accrues. The revenues from policies sold in a given period are called written premium. However, the typical measure of income for a given period is earned premium. The earned premium for a particular policy in a given year is the policy premium multiplied by the}
proportion of the policy term that takes place in that year. Annual industry earned premium is the sum of policy earned premiums and will include policies sold in the current and previous year. For example, half the premium for an annual policy beginning on July 1, 1977, will be included in the 1977 earned premium, as will half the premium of an annual policy beginning July 1, 1976.

The second important feature of SAP concerns treatment of claims costs (acquisition expenses are recorded when they occur). "Incurred loss" is the item closest to estimated claims costs, but it differs from estimated claims costs in several important ways. The incurred loss entry contains an undiscounted estimate of expected losses and loss adjustment expenses associated with premiums earned during the period. In addition, incurred loss for a given year includes revisions to expected losses from previously expired policies. This includes both revisions associated with payments made in the current year and revisions associated with expected future payments for previously expired policies.

- Underwriting margin and specification. The underwriting margin used here is the difference between earned premium and incurred loss as a percent of earned premium: \[ UM_t = \frac{\text{earned premium} - \text{incurred loss}}{\text{earned premium}} \times 100. \] 

I discuss determinants of earned premium and incurred loss separately. The details of the analysis are in Gron (1989, 1990).

Earned premium is determined by insurance pricing and reporting methods. A major cost for insurers is the present value of expected losses. The present value of expected losses depends upon the expected nominal claims payment as well as the expected inflation rate and expected discount rate. In general, the present value of expected losses increases when expected inflation increases, and it decreases when the expected discount rate increases. Under the capacity-constraint hypothesis, price is also negatively correlated with relative capacity.

As illustrated in the accounting example, earned premiums come from a reporting convention that combines revenues from policies priced in the current period as well as policies priced in earlier periods. Thus, if policies cover one-year periods, earned premium for year \( t, EP_t \), will be a weighted average of prices from years \( t \) and \( t - 1 \) (\( P_t \) and \( P_{t-1} \)). In particular, 

\[ EP_t = (1 - \alpha_t)P_t + \alpha_tP_{t-1}, \]

where \( \alpha_t \), the weight for policies priced in year \( t - 1 \), is the fraction of policies sold in \( t - 1 \) that were unexpired at the beginning of period \( t \).

Incurred loss is the other component of the underwriting margin. It can be written as the sum of three variables: unexpected payments during year \( t \) for premiums earned prior to \( t \), revisions to outstanding liabilities for premiums earned prior to year \( t \), and estimated liabilities for premiums earned in period \( t \). The first two components reflect differences between expected payments and actual payments. For example, new information about the expected average loss, random variation such as large idiosyncratic losses associated with a catastrophe, and sampling error will lead to unexpected payments.

Footnote: The insurance literature calls incurred losses divided by earned premium the loss ratio (this includes loss adjustment expenses). The measure used here is \( 1 - \text{loss ratio} \). Acquisition expenses, not included here, are included in the combined ratio, \( \left[ \frac{\text{incurred loss}}{\text{earned premium}} \right] + \left[ \frac{\text{acquisition expense}}{\text{written premium}} \right] \).

Using the combined ratio as the underwriting margin introduces the need to control for acquisition expenses. Market share of direct writers could control for the level of expenses, but it is not available over the period. Basing the underwriting margin on the loss ratio reduces the problem. Some variation in the loss ratio is due to expense changes, but it is assumed the effect is relatively small. The findings of Clarke et al. (1988) support this assumption. They find substantially more variation in the loss ratio than the expense ratio and conclude that "the components of ... insurers' financial statements most responsible for the changes in its position are loss payments and [loss] adjustment expenses" (p. 372).
### Table 1: Definition of Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>Expected Sign</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CAP</strong></td>
<td>Residual from regression of relative capacity on a time trend and constant</td>
<td>&lt;0</td>
</tr>
<tr>
<td><strong>ΔE[INFL]</strong></td>
<td>( E[cost\ inflation]_t - E[cost\ inflation] )</td>
<td>0</td>
</tr>
<tr>
<td><strong>ERRINF</strong></td>
<td>Actual cost inflation, (- E[cost\ inflation] )</td>
<td>&lt;0</td>
</tr>
<tr>
<td><strong>ΔPV</strong></td>
<td>( E[cost\ inflation] - E[nominal\ interest\ rate] )</td>
<td>&gt;0</td>
</tr>
</tbody>
</table>

In addition, for homeowners:
- **CATAS**: The residual from a regression of annual catastrophic losses in thousands of 1983 dollars (those disasters which resulted in greater than $4 million 1983 dollars of damage to insured property) on a constant and a time trend.

The underwriting margin combines both earned premium and incurred loss as described in (1). From the components of earned premium and incurred loss, a reduced-form expression for the underwriting margin (UM) is

\[
UM_t = \beta_0 + \beta_1 CAP_t + \beta_2 CAP_{t-1} + \beta_3 \Delta PV_t + \beta_4 \Delta PV_{t-1} \\
+ \beta_5 ΔE[INFL]_t + \beta_6 ERRINF_t + \beta_7 E[INFL]_t + \epsilon_t.
\]  

(3)

Subscripts denote time periods. **CAP** is relative capacity, **ΔPV** is the change in the net discount rate for expected losses, **ΔE[INFL]** is change in expected inflation, **ERRINF** is error in inflation, and **E[INFL]** is expected inflation. The following describes the variables and their expected effects on the underwriting margin, summarized in Table 1.

**CAP** is relative capacity in period *t*. Under the capacity-constraint hypothesis, variations in relative capacity affect the underwriting margin through short-run premiums. Higher premium levels reduce the ratio of incurred loss to earned premium, increasing underwriting margin. The estimated coefficients on the capacity measures will include the weights \( \alpha \) and \((1 - \alpha)\), where \( \alpha \) is assumed constant over the period estimated.\(^{10}\) The sum of the estimated coefficients on capacity, \( \beta_1 + \beta_2 \), measures the net effect of a change in capacity.

Relative capacity in period *t* may be endogenous. Under the capacity-constraint hypothesis, low relative capacity in period *t* - 1 induces a higher rate of return on net worth in period *t* - 1. Higher underwriting margins, hence retained earnings, and external funds increase period-*t* capacity. Therefore the coefficient on **CAP** may be biased upwards if estimated using ordinary least squares (OLS). Instrumental variables (IV) estimation, however, will be consistent. Exogenous variables that affect the value of insurers' assets are good instruments for **CAP**, as they affect the level of net worth but are not influenced by the industry return to net worth in period *t* - 1. Since the majority of insurer assets are held in stock and bond portfolios, good instruments will be correlated with changes in stock and bond values.

**ΔPV** measures changes in the present value of the expected loss by the net change in the discount factor. The difference between expected inflation and expected nominal interest rates is used to control for changes in the present value of expected losses due to

---

\(^{10}\) If the true \( \alpha \) is not constant, a specification that assumes it is constant may produce biased estimates. The estimated coefficients will include terms of the form \( (\alpha - \alpha_0)X \), where \( \alpha \) is the mean of \( \alpha \) and \( X \) are the independent variables that interact with \( \alpha \). The bias depends upon the correlation between \( (\alpha - \alpha_0)X \) and the independent variables. Whether these variables are correlated and the sign of the possible correlation cannot be determined with the available information.
changes in the discount or expected inflation rate. When this term is larger the present value of expected payments and therefore premiums are greater, \textit{ceteris paribus}. Thus, the coefficient on $\Delta PV$ should be positive. As with relative capacity, the net effect of a change in $\Delta PV$ is also the sum of the coefficients on period $t$ and $t-1$ variables, $\beta_3 + \beta_4$.

Finally, changes in any of the three components of incurred losses will affect the underwriting margin. These are: unexpected payments during year $t$ for premiums earned prior to year $t$, revisions to outstanding liabilities for premiums earned prior to year $t$, and estimated liabilities for year $t$. Unexpected loss payments in period $t$ and revisions to estimated future liabilities are related to the difference between actual period-$t$ inflation and expected period-$t$ inflation, $ERRINF$, as of the end of the previous period. When actual cost inflation in year $t$ is greater than expected, payments in year $t$ will exceed the funds allocated in the loss reserve, increasing year-$t$ incurred losses. Similarly, when losses associated with catastrophic events are higher than expected, claims payments in year $t$ are greater than expected.$^{11}$ Incurred losses also include revisions to expected cost inflation, $\Delta E[INFL]$. Upward revisions to expected future losses increase the nominal value of firms’ outstanding liabilities and increase incurred losses. A third component of incurred losses is expected cost inflation, $E[INFL]$. Higher expected cost inflation increases incurred losses. Higher incurred losses reduce the underwriting margin.

An important simplification has been made concerning the relative timing of premium and loss reserve determination. I assume prices are set at the beginning and loss reserves calculated at the end of the accounting period. If the interval of observation were relatively brief, such as monthly or quarterly, this might not be problematic. The observation interval, however, is annual. Information about the difference between actual and expected inflation revealed during the course of the year may have additional effects not considered here: premiums may increase as expectations of future inflation rates are revised upwards, nominal expected losses for premiums earned in period $t$ may be higher (through greater expected inflation), and premiums may increase if the unanticipated payments and revisions to loss reserves substantially reduce capacity. To the extent that these feedback effects are significant, this specification will only capture the net annual effect of these changes on accounting profitability.

\section*{Data.} The analysis uses aggregate stock insurer time series of underwriting margins for automobile liability, auto physical damage, homeowners’ multiple peril, and other liability insurance from 1952 to 1986. Automobile liability, automobile physical damage, homeowners’ multiple peril, and other liability are the four largest lines of insurance in terms of revenues. Automobile insurance accounts for about 40% of total property-casualty sales, with auto liability being 20–25% of the total and auto physical damage being 14–18%. Homeowners’ multiple peril was introduced in the early 1950s and did not have substantial market share until the early 1960s. From 1960 to 1986, homeowners’ revenues were 8–12% of total insurance sales. The fourth line, other liability, made up 6–7% of total revenues from 1952 to 1975. From then on, other liability’s share of total insurance revenues was in the general range of 9–15%.$^{12}$ The sample is restricted to stock insurers (70–75\% of the industry during the period) to control for the cost of generating capital. The regulation of mutual insurers, the other major ownership form, is likely to give them a higher cost of capital. The restriction to stock insurers is unlikely to bias the results, since the omitted group has the same or higher capital costs and has substantially smaller market share (20–30\%).

\footnotetext[11]{The costs of catastrophic events are reported in Insurance Information Institute (annual). These data are for property losses only, so they apply only to homeowners’ insurance.}

\footnotetext[12]{Data are from Best’s Aggregates and Averages. Data for homeowners’ insurance begins in 1957 because little was sold earlier. Other liability is all commercial liability other than commercial multiple peril, workers’ compensation, and commercial auto. It includes medical malpractice.}
Figure 2 is a plot of each time series. As this shows, underwriting margins are not perfectly correlated across lines. Some of the variation comes from differences in timing of payments. The time between the accident event and actual payment, known as the "tail" of the line, differs across lines. Of the four lines examined here, auto physical damage and homeowners' multiple peril have relatively short payment periods; for example, from 1977 to 1986, over 95% of homeowners' claims were paid within three years after the policy period. Auto liability and other liability have significantly longer tails; in the period 1977–1986 it took 5 to 6 years for 95% of auto liability claims to be paid, and 9 to 10 years for other liability claims. Since the underwriting margin does not discount expected losses, increases in expected inflation and interest rates will have larger effects on lines with longer tails. This is evident in the greater underwriting margin decline for the two liability lines in Figure 2.

Differences in cost and production factors will also cause underwriting margins to vary by line. For example, automobile repair costs have a large influence on the cost of auto physical damage coverage, while medical care costs do not. On the other hand, the cost of medical care is important for other liability costs. Underwriting margins will also vary across lines if firms cannot costlessly shift capacity between lines.

The correlation between the all-lines underwriting margin and that for each line is as follows: auto liability (.96), auto physical damage (.80), homeowners' (.72), and other liability (.88). The other liability series looks very different before the late 1960s. The correlation between the all-lines and the other liability underwriting margin is only .10 before 1968, while it is .91 for the post-1968 period.

Figures are from the author's calculations of Schedule P data for 25 insurer groups. Auto physical damage is not in Schedule P, as it is not a long-tailed line. It is assumed that the tail for auto physical damage is no greater than that of homeowners' (data are from Best's Convention Statements).

There may be a cost to reducing net worth allocated to a specific line of insurance. For example, if buyers face search costs, it may be costly for the firm to reduce current quantity, since this will affect future output when the rate of return for that line of insurance may be higher.
Capacity by line is not available because insurers do not report net worth allocated by line of insurance.\textsuperscript{16} Capacity is measured as the residual from a regression of relative net worth on variables related to equilibrium industry capacity. Relative net worth in year \( t \) is aggregate stock insurers' policyholders' surplus as of December 31, year \( t - 1 \), divided by year-\( t \) gross national product (GNP).\textsuperscript{17} It is the level of net worth the industry has upon entering the new period relative to the quantity of goods and services that can be insured during that period. I have used two capacity variables. The measure reported in the text is the residual from regressing relative net worth on a constant and a time trend. Results using a second measure, which controls for changes in economic activity using a measure of the GNP gap, are reported in Appendix B.\textsuperscript{18} Although the second measure is available for a shorter time period, the results are similar.

The remaining variables were constructed using inflation and interest rate series. Inflation measures vary by line. The medical care cost index from the consumer price index was used to construct the inflation measures for auto liability and other liability. The automobile maintenance and repair index was used for auto physical damage, while the consumer price index was used for homeowners' multiple peril. To calculate expected inflation in year \( t \), I estimate the following equation using data up to year \( t - 1 \).

\[
I_s = \alpha_0 + \gamma_1 I_{s-1} + \gamma_2 I_{s-2} + \gamma_3 I_{s-3} + \nu_s, \tag{4}
\]

where \( I_s \) is the annual inflation rate for year \( s \). Expected inflation for year \( t \) is then \( I_t \), the fitted value from (4). Expected nominal interest rates are determined in the same manner, using the interest rate on medium-term government bonds.\textsuperscript{19} Errors in expected inflation and nominal interest rates are computed as the difference between actual and expected inflation.\textsuperscript{20}

The relevant horizon for forming expectations is from the time new information becomes known to the time all outstanding liabilities are expected to be settled. I assume the net effect of all expectations and changes in expectations can be approximated by a model with a one-period horizon. The expectations are described in Table 1. This simplification is likely to be more restrictive for longer-tailed lines than for shorter-tailed lines.

\textsuperscript{16} Some instruments available for relative capacity do vary by line. Measures of demand by line are very poor. The best measures exist for auto insurance, where we can estimate the number of cars. However, demand for auto insurance depends also on the value of cars (and income levels, etc.). Furthermore, estimates of number of cars are based upon automobile registrations, which vary with insurance prices and enforcement in states with mandatory insurance. For other lines of insurance, a measure of potential demand is even more problematic.

\textsuperscript{17} Capacity measures including other companies' net worth are very similar. The correlation between \{stock net worth / GNP\} and \{stock + mutual net worth / GNP\} is .97; that between \{total industry net worth / GNP\} and \{stock net worth / GNP\} is .95. Using total industry net worth instead of stock insurer net worth does not change the conclusions.

\textsuperscript{18} Net written premiums divided by net worth is often cited in discussions of industry capacity. It measures "a company's exposure to pricing errors in its current book of business" (A.M. Best Company, 1985). If the ratio is two and premiums turn out to be 10% below actual costs, then 20% of net worth is needed for payments. This interpretation requires expected costs to be a constant fraction of premiums; if the fraction varies, the meaning of this measure is unclear. In addition, this measure of capacity is endogenous under any theory of insurance cycles. Net written premiums are highly correlated with earned premiums.

\textsuperscript{19} This methodology is used by Barro and Sala-i-Martin (1990) to estimate expected inflation. Gordon and Veitch (1986) use a similar procedure. For the interest rate series, three-year to five-year bonds are used prior to 1952, five-year bonds after that.

\textsuperscript{20} I assume premiums are priced at the beginning of the period, so they are based on information known at the beginning of year \( t \) (up through \( t - 1 \)). Since incurred losses and loss reserves are from end-of-year annual statements, insurers calculate these measures using information known at the end of the year. Therefore, the relevant change in expected inflation is the expected inflation next period \( t + 1 \) less the expected inflation this period \( t \), while the difference between expected inflation and expected nominal interest in year \( t \) is the measure that affects the present value of losses in year \( t \) used in determining premiums.
4. Estimation results

Results from ordinary least squares (OLS) and instrumental variables (IV) estimation of (3) are reported in Table 2. The standard errors reported in Table 2 are robust to the presence of serially correlated errors as well as heteroskedastic errors.\(^{21}\) The results for lines with relatively short loss tails support the capacity-constraint hypothesis. OLS results for both lines indicate that lagged capacity variables have a statistically significant negative effect on accounting profitability. Period-\(t\) capacity has a large, positive effect on the auto physical damage underwriting margin significant at the 30% level (in a two-tailed test), although the total effect of capacity is negative and significant above the 2.5% level (in a one-tailed \(t\) test). Under the capacity-constraint hypothesis, period-\(t\) capacity (and period \(t - 1\) for homeowners’ multiple peril) is endogenous, and the coefficient will be biased upwards. IV estimation substantially reduces both the magnitude and the significance of the period-\(t\) capacity variable in the auto physical damage equation.

OLS results for auto liability accounting profits are less significant, but they also provide results consistent with the capacity-constraint hypothesis. The results improve substantially with IV estimation. The magnitude of the estimated effect of capacity increases from \(-.04\) to \(-.08\) and is statistically significant at the 5% level (in a one-tailed test).

The results for other liability support neither the capacity-constraint hypothesis nor arbitrage theory. The industry capacity variable has a positive, statistically significant effect on accounting profitability.

<table>
<thead>
<tr>
<th>TABLE 2</th>
<th>Regression Results</th>
<th>Underwriting Margin for Individual Line of Insurance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient on</td>
<td>Auto Physical Damage</td>
<td>Homeowners’ Multiple Peril</td>
</tr>
<tr>
<td>OLS</td>
<td>IV</td>
<td>OLS</td>
</tr>
<tr>
<td>(CAP)</td>
<td>-.11</td>
<td>-.14</td>
</tr>
<tr>
<td>(.05)</td>
<td>(.05)</td>
<td>(.05)</td>
</tr>
<tr>
<td>(\Delta E[INF,])</td>
<td>-.10</td>
<td>-.129</td>
</tr>
<tr>
<td>(.88)</td>
<td>(.93)</td>
<td>(1.28)</td>
</tr>
<tr>
<td>(ERR INF)</td>
<td>-.41</td>
<td>-.48</td>
</tr>
<tr>
<td>(.54)</td>
<td>(.55)</td>
<td>(.86)</td>
</tr>
<tr>
<td>(E[INF,])</td>
<td>-.260</td>
<td>-.276</td>
</tr>
<tr>
<td>(.49)</td>
<td>(.57)</td>
<td>(.36)</td>
</tr>
<tr>
<td>(\Delta PV)</td>
<td>.95</td>
<td>1.00</td>
</tr>
<tr>
<td>(.23)</td>
<td>(.24)</td>
<td>(.32)</td>
</tr>
<tr>
<td>(CATAS,)</td>
<td>-3.2E-6</td>
<td>-3.1E-6</td>
</tr>
<tr>
<td>(N)</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>(R^2)</td>
<td>.52</td>
<td>.50</td>
</tr>
<tr>
<td>SEE</td>
<td>5.02</td>
<td>5.09</td>
</tr>
</tbody>
</table>

Regression done using robust standard errors correction.


\(^{21}\) The Durbin-Watson statistics for the regressions ranged from 1.10 to 1.60. Estimation uses the Newey and West (1987) procedure with one lag to correct for serial correlation and heteroskedasticity.

Estimation for homeowners’ multiple peril includes \(CATAS,\), a measure of catastrophic losses. Some homeowners’ policies are written for periods longer than a year, so longer lags of some variables may be significant. \(CAP\) and \(\Delta PV\) from period \(t - 2\) were significant. All results for homeowners’ multiple peril reported here include two lags of these variables.
fect on accounting profit. While IV estimation alters the coefficient on $CAP$, in a direction consistent with the capacity-constraint theory, the net effect of capacity on the other liability underwriting margin is still positive.

There are several reasons why this specification is less successful with other liability. First, it is more difficult to approximate movements in other liability capacity with industry capacity. Other liability is the smallest of the four in terms of revenue share, and reinsurance capacity is particularly important for other liability but is not included in capacity because no reliable data exist. Second, other liability insurance changed substantially over the period 1952–1986. In the first half of this period, large tort awards were rare and mass toxic torts did not exist. By the end of the 1960s, however, many significant changes in the tort system were in place, resulting in large tort awards. An $F$ test for change in regime between the periods 1952–1968 and 1969–1986 rejected the null hypothesis of no change at the 5% level (see Appendix B). The significant, positive coefficient on capacity in the 1969–1986 period suggests this variable may be capturing increases in demand as well as supply effects. Lastly, it is particularly difficult to control for movements in other liability incurred losses. For example, the emergence of mass toxic torts and catastrophes such as the 1981 collapse of the Hyatt Hotel walkway create large changes in incurred losses. The data to control for these types of events are also unavailable. For these reasons, the results for other liability will be omitted from the remaining tables and left for future research.

A test for the endogeneity of $CAP$, was performed using a regression form of a Lagrange multiplier test robust to serial correlation and heteroskedasticity (Wooldridge, 1989). The test fails to reject the null hypothesis, that $CAP$, is exogenous, at the 10% level for auto physical damage and homeowners’ multiple peril. For auto liability the null hypothesis was rejected at the 2.5% level. The feedback effects between profitability and industry capacity are empirically less significant for auto physical damage and homeowners’ multiple peril than for the longer-tailed line, auto liability.

The other estimated coefficients are largely consistent with expectations and significantly different from zero at the 10% level. An exception is the change in expected inflation, which should have a negative coefficient since higher expected inflation causes positive revisions to the loss reserve, $ceteris paribus$. Policies are written throughout the year, however, and increases in expected inflation also increase premiums by increasing the present value of expected losses. It appears that for longer-tailed lines the premium effect dominates the increase in incurred losses. The coefficient on $\Delta E[NFLX]$, may also be capturing capacity effects: large increases in expected inflation will cause adjustments to the loss reserves, reducing net worth. Longer-tailed lines have a larger proportion of claims outstanding, so revisions may have a larger impact on capacity.

A comparison of coefficients across lines lends confidence to the results, since non-capacity coefficients have signs and relative magnitudes consistent with predictions based on regulatory and accounting rules. $\Delta PV$ has a larger and more significant effect on lines with longer tails. Some of the greater volatility observed in accounting profitability of longer-tailed lines in the latter half of the period can be attributed to greater swings in

---

22 Winter (1991) suggests that changes in the reinsurance market during the early 1980s produced a positive relationship between the underwriting margin and capacity as measured here.

23 114 people were killed and 200 injured when a walkway collapsed at the Hyatt Regency Hotel in Kansas City. Over 300 lawsuits and other liability claims were filed, according to Insurance Information Institute (annual).

24 A test on other liability rejected the null hypothesis at just over 5%.

25 The test was repeated using a nonrobust form of the test statistic, as the robust form may be less likely to reject the null hypothesis. The outcome was the same. Since auto liability accounts for the largest share of premium volume, the capacity measure used here may be a better measure of auto liability capacity than of the others.
interest and inflation rates, which have a larger effect on the underwriting margin of these lines. For example, an increase of 1 point in $\Delta PV$ increases the loss ratio of auto physical damage by 1 point and the loss ratio of auto liability by 2.4 points.\textsuperscript{26}

The estimation procedure thus far assumes that insurers fully adjust loss reserves when new information is learned and that loss reserves are not discounted; that is, it assumes that insurers truthfully report financial statistics each period.\textsuperscript{27} Insurers may, however, find it optimal to smooth underwriting profit. Firms may understate incurred losses when income is low and overstate them when income is high.\textsuperscript{28}

Smoothing is most likely to occur in the longer-tailed lines. Reserves for these lines cover longer periods; there is more uncertainty as to the final payment and more room for manipulation. If insurers partially adjust loss reserves as new information is revealed, lagged values of inflation errors and changes in expected inflation should have a negative effect on the underwriting margin. As claims are settled and paid in subsequent periods, insurers will recognize "old" news. Errors in inflation will have a negative effect on the underwriting margin in the next period if insurers are smoothing underwriting results. However, under the capacity hypothesis, large errors in inflation expectations will also lead to adjustments in reserves that reduce net worth. This implies a positive effect of past errors in expectations: lower net worth reduces capacity and increases the next period's profitability.

Table 3 reports the results from (3), including lagged errors in inflation, to test whether insurers smooth reported profits. For auto liability, in the last two columns, the positive coefficient on lagged errors in expected inflation supports the capacity hypothesis. Any smoothing effects are dominated by the effect on capacity of unanticipated losses. The inclusion of lagged errors decreases the magnitude and significance of the capacity variable, indicating that it may measure some of the same effects. The net effect of errors in inflation expectations on accounting profit is negative: the negative effect of current un-expected losses is substantially larger than the positive effect of the lagged variable. When lagged errors are included in the auto physical damage and homeowners' multiple peril equations, the estimated effects are smaller and not statistically significant. In the auto physical damage equation, the coefficient on lagged errors in expected inflation is negative, and its inclusion reduces the coefficient on change in expected inflation. The coefficient on lagged error in expected inflation is positive in the homeowners' multiple peril equation, and its presence increases the estimated effect of both period-$t$ error in expected inflation and change in expected inflation. The sum of the coefficients on the error in expected inflation is approximately the same as the estimated effect of period-$t$ error from Table 2.

The results show that variations in capacity affect insurer profitability in a manner consistent with the capacity-constraint hypothesis when other factors are controlled for. For short-tailed lines, which are predominantly property coverages, the industry measure of variations in capacity approximates by-line capacity reasonably well. A robust test for endogeneity of contemporaneous capacity failed to reject the null hypothesis of exogeneity for auto physical damage and homeowners' multiple peril, but it rejects the null hypothesis for auto liability and other liability. For automobile liability, the industry capacity variable

\textsuperscript{26} Including errors from (4) to measure unexpected cost inflation ($ERRINF$) will bias the estimated coefficient on $ERRINF$ if there are omitted variables that are correlated with $ERRINF$, such as unexpected inflation in other costs. To the extent that $ERRINF$ is included to control for all expected cost inflation, this is not a significant problem, as the coefficient is not being used to predict the effects of changes in a specific cost on the underwriting margin. If $ERRINF$ is orthogonal to the other variables in (3), correlation with the error term will not bias other coefficients. This suggests care should be exercised in interpreting the coefficients in Table 2.

\textsuperscript{27} In general, discounting of loss reserves was not allowed by statutory accounting principles prior to the Tax Reform Act of 1986. The exception, worker's compensation, is not considered here (Strain, 1986).

\textsuperscript{28} Weiss (1985) finds evidence consistent with income smoothing.
had a smaller effect on profitability, but the effects of other variables on the underwriting margin are consistent with the capacity-constraint story. In particular, the effects of lagged inflation errors and the effect of $\Delta E[INFL]$ provide support for the role of insurer capacity in automobile liability premium determination.

5. Summary and conclusion

The regression analysis demonstrates that variations in capacity have a significant, negative effect on movements in property-casualty profitability. This supports the capacity-constraint hypothesis where the short-run industry supply curve is upward sloping. The finding applies to two areas of insurance policy: price regulation and guaranty fund operation.

Arbitrage models of insurance pricing underlie most policy discussions of the industry, and analysts often use them as important criteria for determining whether insurance markets operate competitively. The results from this article indicate that actual insurance prices are likely to diverge from those predicted by arbitrage models even if insurers behave competitively. Therefore, regulators should develop additional criteria to use in judging insurance market operations.

Another implication of these results is that the current form of most guaranty funds may exacerbate the effects of a capacity shortage. Insurance buyers are protected from insurer insolvency through state guaranty funds. These "funds" generally operate as follows. Once an insurer is declared insolvent, the state insurance department assesses all licensed insurers (in that state) to contribute to the fund. Although there is a maximum

---

29 Only New York's guaranty fund operates with a prefunded system.
annual assessment that can be made, insurers may be assessed several years in a row. Policyholders of the insolvent insurer will have their claims paid up to a state-set maximum—usually $300,000.30

Rules governing how the fund operates vary by state. Guaranty funds differ in the lines of insurance they cover and the limit per claim. Some states separate insurance into accounts by line and assess only those insurers licensed to write the same lines as the insolvent company. Many states provide for a mechanism for insurers to police one another. Organizations called state guaranty associations, made up of insurers licensed in the state, were intended to prepare reports on the causes of insolvencies and “alert the regulators about financially troubled companies.” The associations have not operated in this manner, owing to potential antitrust allegations.

If underwriting cycles are associated with variations in capacity, this form of guaranty funds will exaggerate any capacity shortages. Gron (1989, 1990) shows that large numbers of insolvencies are correlated with the “crisis” phase of the cycle. A regulation that further reduces net worth by assessing solvent insurers will further reduce industry capacity. While guaranty funds do protect a few consumers from insurer insolvencies, they may increase premium level throughout the industry by reducing net worth at a time when the industry is already capacity constrained. A prefunded system, such as New York’s, avoids these problems, but a full analysis of the two systems is left for future research.

Appendix A

- Accounting example. It is well known that standard accounting measures differ from the desired economic variables. Property-casualty insurance accounting rules, known as statutory accounting practices (SAP), differ substantially from usual accounting procedures known as generally accepted accounting practices (GAAP), and these differences affect the variables used in this analysis. This example illustrates how cash flows correspond to SAP income and balance sheet entries.

Consider the two-year (y1 and y2) experience of an insurer that writes one auto insurance policy each year. The policy term is 12 months; the insurer expects to pay out $80 at an average of 16 months after the policy is sold. With a 10% annual discount rate and monthly compounding, the present value of the expected loss is $70 at the time the policy is sold. Acquisition expenses of $20 are paid at the time the policy is sold, and acquisition costs plus discounted expected payments include a competitive return for the insurer. The insurer starts out at the beginning of y1 with $500 capital, or policyholders’ surplus. Insurer assets are invested in instruments that earn a 10% annual interest rate compounded monthly.

On July 1, y1, the insurer sells an auto policy effective from that date until June 30, y2. The buyer pays the competitive premium of $90 on July 1. In y2, the policy is renewed. This second policy covers July 1, y2 to June 30, the following year. The premium is again $90. In November of y2, 16 months after the first policy was sold, a claim is made on the first policy. The realized loss is $95, $15 more than the expected value. In December of y2, the insurer revises the expected payment on the second policy to $70 because of greater enforcement of drunk driving laws and speed limits.

How do these cash flows translate into property-casualty insurance accounting numbers? Table A1 shows the income and balance sheets for the insurer. Property-casualty insurers typically break down income statements between underwriting and investment activities. The underwriting income statement involves premium and payment accounts, while the investment statement records income from financial assets. I will follow the same convention.

SAP recognize premium revenue as income only as it accrues. Therefore, when the first policy is initially sold, none of the $90 premium is recorded as income, since none of the policy period has expired. Instead, a liability called the unearned premium reserve (UPR) is created. As the policy term elapses, the premium is said to be “earned.” A portion of the premium corresponding to the elapsed policy period is deducted from the UPR and accounted as income under the heading “earned premium.” At the end of y1, half the policy has expired. The earned premium in the income statement shows $45, and on the balance sheet the UPR shows the remaining $45.31 In y2, the insurer earns the remaining $45 of the first premium and sells the second policy

30 The primary source for this discussion is U.S. General Accounting Office (1987).
31 The UPR shows that until the full policy term expires, the buyer can cancel the remaining portion of the policy and the insurer must refund a portion of the premium. Acquisition costs have not been deducted from the UPR, so the UPR will overestimate the potential repayment the insurer would have to make, and the earned premium will underestimate the actual revenue the insurer has “earned.”
### TABLE A1

#### Year 1 (y1)

<table>
<thead>
<tr>
<th>Income Sheet</th>
<th>Investment Income</th>
<th>Balance Sheet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earned premium</td>
<td>Return for the year on $500</td>
<td>Assets</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>52.36</td>
</tr>
<tr>
<td>Incurred loss</td>
<td>Return for 6 months on $70</td>
<td>3.57</td>
</tr>
<tr>
<td></td>
<td>(40)</td>
<td>625.93</td>
</tr>
<tr>
<td>Acquisition expenses</td>
<td>(20)</td>
<td></td>
</tr>
<tr>
<td>Net underwriting income</td>
<td>(15)</td>
<td>Net investment income</td>
</tr>
<tr>
<td></td>
<td></td>
<td>55.93</td>
</tr>
</tbody>
</table>

#### Year 2 (y2)

<table>
<thead>
<tr>
<th>Income Sheet</th>
<th>Investment Income</th>
<th>Balance Sheet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earned premium</td>
<td>Return for 10 months on $625.93</td>
<td>54.18</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>667.57</td>
</tr>
<tr>
<td>Incurred loss</td>
<td>Return for 2 months on $530.93</td>
<td>8.89</td>
</tr>
<tr>
<td></td>
<td>(90)</td>
<td></td>
</tr>
<tr>
<td>Acquisition expenses</td>
<td>Return for 7 months on $70</td>
<td>3.57</td>
</tr>
<tr>
<td></td>
<td>(20)</td>
<td></td>
</tr>
<tr>
<td>Net Underwriting Income</td>
<td>(20)</td>
<td>Net investment income</td>
</tr>
<tr>
<td></td>
<td></td>
<td>66.64</td>
</tr>
</tbody>
</table>
for $90. The earned premium entry at the end of y2 reflects both the second half of the first policy and the first half of the second policy. The UPR indicates that half the second policy term is unexpired.

Costs are also measured using SAP. Acquisition expenses are recorded when they occur. Loss payments are dealt with in a more complex manner. A liability known as the loss reserve measures the current undiscounted estimate of all outstanding claims liabilities at the end of the year. At the end of y1, the loss reserve reflects that half of the first policy has expired. Assuming the insurer expects to pay claims equal to one-half the total expected loss for a period covering half the full term, the loss reserve is $40 \( (1/2 \times 80) \).

On the income statement the accounts associated with loss payments are reported under the heading “incurred loss.” Incurred losses are the change in the loss reserve, plus losses paid. For y1 this is the same as the loss reserve, since the loss reserve began the year at zero and no claims were paid. In y2 three features make calculating the loss reserve and incurred losses more complicated: a revision to the expected loss from policy two, a payment made on policy one, and an upward adjustment for the difference between the expected and actual payment on policy one. At the end of y2 the loss reserve is $35 (one-half the expected loss from policy two). Incurred losses are the change in the loss reserve plus losses paid \( ((35 - 40) + 95) = 90 \). The incurred loss entry reflects not only expected losses for policies in effect in y2, but also revisions to expected losses on elapsed or future policies made in y2.

Calculation of the investment income is straightforward. It is the total funds available to the insurer compounded monthly using a 10% annual interest rate.

Appendix B

<table>
<thead>
<tr>
<th>TABLE B1 Other Liability Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAP_1</td>
</tr>
<tr>
<td>CAP_{t-1}</td>
</tr>
<tr>
<td>ΔE[INFL]_t</td>
</tr>
<tr>
<td>ERRINF_t</td>
</tr>
<tr>
<td>E[INFL]_t</td>
</tr>
<tr>
<td>ΔPV_{t-1}</td>
</tr>
<tr>
<td>ΔPV_{t+1}</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>R^2</td>
</tr>
<tr>
<td>SEE</td>
</tr>
</tbody>
</table>


Another way to calculate the loss reserve in y2 is to make the appropriate adjustments to the loss reserve at the end of y1. As described in the following section, the loss reserve as of y2 is equal to the loss reserve at the end of y1 plus expected payments for premiums earned in y2 net of payments made during y2 plus revisions to expected future payments on premiums earned prior to y2 minus the reserves set aside at the end of y1 to pay for claims paid in y2. In this case the loss reserve at the end of y2 is

\[ 40 + (40 + 35 - 40) + 0 - 40 = 35. \]
### TABLE B2
Regression Results Using Other Capacity Measure
Dependent Variable: Underwriting Margin for the Line

<table>
<thead>
<tr>
<th>Coefficient on</th>
<th>Auto Physical Damage</th>
<th>Homeowners' Multiple Peril</th>
<th>Auto Liability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>IV</td>
<td>OLS</td>
</tr>
<tr>
<td>CAP</td>
<td>-0.07</td>
<td>-0.10</td>
<td>-0.16</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.07)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>ΔE[INFL] t</td>
<td>-3.93</td>
<td>-4.34</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>(1.84)</td>
<td>(1.91)</td>
<td>(0.98)</td>
</tr>
<tr>
<td>ERRINF,</td>
<td>1.14</td>
<td>1.15</td>
<td>-0.56</td>
</tr>
<tr>
<td></td>
<td>(0.86)</td>
<td>(0.88)</td>
<td>(0.59)</td>
</tr>
<tr>
<td>E[INFL] t</td>
<td>-1.73</td>
<td>-1.85</td>
<td>-1.21</td>
</tr>
<tr>
<td></td>
<td>(0.48)</td>
<td>(0.50)</td>
<td>(0.33)</td>
</tr>
<tr>
<td>ΔPV</td>
<td>0.39</td>
<td>0.36</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>(0.43)</td>
<td>(0.44)</td>
<td>(0.32)</td>
</tr>
<tr>
<td>CATAS,</td>
<td>-3.1E-6</td>
<td>-3.3E-6</td>
<td>2.90</td>
</tr>
<tr>
<td></td>
<td>(1.4E-6)</td>
<td>(1.45E-6)</td>
<td>(0.72)</td>
</tr>
</tbody>
</table>


### References


