

Diversification as a Process: Geographic Expansion and the Formation of National Insurers*

Pawel Janas[†]

This version: October, 2026

Abstract

Life insurance was the largest contractual savings vehicle for American households in the late nineteenth and early twentieth centuries, and between 1888 and 1940 its providers expanded from regional companies into national intermediaries. This paper documents that geographic expansion, using newly digitized data on the premiums, claims, and insurance in force of the near-universe of legal-reserve life insurers across states, and examines whether it made insurers more stable. Insurers became national by entering states incrementally: entered markets opened at small scale, their returns were highly skewed and difficult to predict, and by a firm's twentieth year they supplied most of its premium. Because insurers grew by entering markets, the broadest insurers were also the largest. Conditional on size, geographic breadth did not make insurers more stable. Insurers pooled lives, not places: their stability came from the number of lives they insured, because the claims of insured lives were only weakly correlated within states. Geographic expansion was nonetheless costly. Before 1920, insurers that had expanded across many states paid claims per dollar of insurance in force about 19 percent higher than insurers of the same size writing in the same markets, a firm-wide difference consistent with limited managerial attention to the selection of risks. The 1918–1919 influenza pandemic, a large and geographically uneven mortality shock, raised claims where it was severe, but differences in insurers' geographic exposure account for little of the variation in their claims.

JEL Classification: G22, G23, N21, N22

Keywords: financial intermediation, insurance, risk pooling, geographic diversification

*I thank the Linde Institute at Caltech and the Economic History Association (Cole Grant) for generous financial support. I have benefited greatly from conversations with Tracy Dennison, Phil Hoffman, Sarah Quincy, Jean-Laurent Rosenthal, Andrew Sinclair, as well as seminar participants at Caltech, Harvard, Yale, and the WEAI, ASSA, and SEA annual meetings.

[†]Caltech and NBER, email: pawel@caltech.edu, site: paweljanas.com

1 Introduction

Between 1888 and 1940, American life insurance was transformed from a regional industry into a national one (Janas, 2026). The number of legal-reserve life insurers rose from about 50 to over 350, premium income rose from 0.7 percent of GDP in 1890 to 4.3 percent in 1935, and an industry whose business had been concentrated in the Northeast came to be organized around companies that wrote policies in nearly every state. Life insurers were the period’s largest nonbank financial intermediaries and, through their investments in mortgages and municipal and corporate bonds, a principal source of long-term credit (Davis, 1965; Snowden, 1995). The first purpose of this paper is to document how this intermediary became national by using a newly constructed long-run database of firm-level market outcomes. I digitize the premiums, claims, new insurance written, and insurance in force of every legal-reserve life insurer in every state in which it operated, annually from 1888 to 1940, and link these market-level data to the insurers’ own company accounts and to the statements they filed in New York. The data describe where each insurer sold policies, when it entered and left each state, and how its policies performed in each market.

The second purpose is to use this expansion to examine the stability of financial intermediaries. Theories of financial intermediation locate the advantage of a large intermediary in diversification: an intermediary that holds many imperfectly correlated risks bears less risk per claim it issues (Diamond, 1984; Boyd and Prescott, 1986). Size and geographic breadth are difficult to distinguish in this respect, however, because an intermediary becomes diversified by growing and grows by entering new markets. I ask whether the largest insurers were more stable because they were large or because they were geographically diversified, and what geographic expansion cost the insurers that undertook it. An insurer became broad only by entering markets, so any cost of breadth is a cost of geographic expansion, and the relevant comparison is between insurers of the same size that grew by entering many states and insurers that grew within a few.

The closest evidence on these questions concerns the banking sector. Demsetz and Strahan (1997) find that large bank holding companies are better diversified than small ones, but that better diversification does not translate into lower risk, because large holding companies hold less capital and riskier loans. Goetz et al. (2016) find that the geographic expansion of bank holding companies reduces their risk by reducing their exposure to idiosyncratic local risks, particularly when they expand into economically dissimilar markets. Carlson and Mitchener (2006) find that branch banking stabilized state banking systems in the 1920s and 1930s chiefly by increasing competition and removing weak banks, rather than by diversifying the portfolios of the branch banks themselves.

Banks and insurers perform similar functions in the relevant respects: each issues claims to households (deposits or policies), against a portfolio of risks (loans or insured lives), and each is stable to the extent that the realized losses on that portfolio are predictable. The theory of diversification within an intermediary (Diamond, 1984) applies to both. The banking evidence has been shaped, however, by the institutions that generated it: geographic expansion was, for much of its history, governed by branching law, the location of the risks being diversified must be inferred from the location of branches or deposits, and the possibility of runs affects the relation between diversification and failure.

Life insurance adds to this evidence in four respects. First, the risk an insurer diversifies is the mortality of the lives it insures, and those lives reside where the policies were sold; because the data record each insurer's premiums, claims, and insurance in force state by state, the geographic distribution of the insured risk is observed directly rather than inferred. Second, an insurer could enter any state that licensed it, so the pace and extent of its expansion reflect the firm's choices rather than branching law, and for most insurers the data observe that expansion from the firm's first years. Third, the period precedes state guaranty associations and federal support for insurers, and life insurance liabilities were not subject to runs in the manner of demand deposits, so the relation between breadth and stability reflects underwriting outcomes rather than public guarantees or the coordination of creditors. Fourth, because claims are observed in each market, the claims experience of the markets an insurer entered and that of the markets it already served can be measured separately within the same firm, so that a cost of expansion can be attributed either to the markets entered or to the firm as a whole.

The first set of findings describes the expansion. Insurers became national through many small entries rather than through acquisitions of large positions. Market entries rose from 182 in the 1890s to 958 in the 1920s, and about seven of every ten were into Southern and Midwestern states. Premium migrated from the markets in which an insurer was founded to those it entered: founding states supplied 68 percent of premium at firm age 5 and 43 percent at age 20, and an entered position's premium grew more than threefold, relative to the rest of the firm's business, within a decade. The returns to entry were highly skewed: about a fifth of entries were abandoned within a decade, the best tenth of entries produced about half of a cohort's ten-year premium, and information available at entry predicts almost none of this variation. Entered markets were written at lower prices, because a new entrant held a small local share, and had lower initial loss ratios, because newly selected lives incur few claims. Most insurers nevertheless remained regional:

of the insurers first appearing in the 1900s and 1910s, about a third reached three effective states within a decade and about one in twenty reached ten. The industry's growth therefore combined the expansion of a few national insurers with the formation of many local ones, so that throughout the period insurers headquartered in other states wrote about nine tenths of the premium in the average state, while after 1905 the median insurer collected most of its premium in its home state. Because insurers grew by entering markets, breadth and size moved together, and firm-size deciles explain 57 percent of the cross-sectional variation in breadth.

The second set of findings concerns the cost of expansion. Before 1920, an insurer one standard deviation broader than another of the same size paid claims per dollar of insurance in force 19 percent higher. To determine whether this difference reflects the markets that broad insurers entered or the insurers themselves, I decompose each insurer's aggregate outcomes exactly into a firm-wide level, the markets in which it wrote, and its positions within those markets. The claims differential lies entirely in the firm-wide level: a broad insurer paid more claims than a narrow insurer of the same size in the same markets and positions, and among insurers writing in the same state and year, claims rose with breadth by 27 percent of the mean before 1906 and 17 percent in 1906–1920. The differential did not arise as a discrete deterioration at the time of expansion: in an event study of the markets an insurer already served, claims did not rise after an episode in which it entered three or more states. The differential appears in the death claims and the number of deaths reported in the insurers' own accounts, it is not explained by policy mix or participating business, and it was present throughout a broad insurer's business, including its headquarters state. It declined from about 1915, as the selection of risks was standardized across the industry, and had disappeared by the late 1920s. Its incidence and timing are consistent with limited managerial attention to the selection of risks, under which an insurer that operated in many markets selected the lives it insured less effectively in all of them, although the absence of a discrete change at expansion indicates that the cost was associated with the breadth an insurer had attained rather than with the act of entering markets. The pricing cost of expansion, by contrast, lay in the composition of the firm's business, in the small local shares it held in entered markets, and amounted to about 1.5 percent of the premium those markets earned.

The third set of findings concerns stability. Broad insurers were less volatile and survived longer than narrow ones, but neither difference survives conditioning on size. Decomposing each insurer's loss-ratio variance shows why. The variance that arises from the number of claims in a small portfolio falls with the number of lives insured, wherever they reside; it accounts for nearly two thirds

of the variance of small, narrow insurers and less than a tenth for large, broad ones. The variance that geographic diversification could remove, that arising from mortality fluctuations specific to the states in which an insurer wrote, accounts for about a tenth of the remaining variance in the median insurer. For breadth to have reduced that remaining variance by a fifth, the geographic share would have had to be more than three times as large. Insurers pooled lives, not places, because the claims of insured lives were only weakly correlated within states. The contrast with banks, whose borrowers in a market depend on the same local economy, follows from the difference in that correlation.

The 1918–1919 influenza pandemic provides a test of this interpretation. The pandemic was the largest mortality shock of the period, and its severity varied substantially across states for reasons unrelated to insurers’ prior geographic choices. At the market level, it raised claims where excess mortality was higher: within an insurer, a market’s loss ratio rose by 0.018 for each additional excess death per 1,000 inhabitants. Breadth reduced the dispersion of insurers’ exposure to the pandemic by more than half, but the dispersion it removed was small relative to other sources of variation in insurers’ claims. Differences in exposure account for 0.4 percent of the variation in the change in insurers’ loss ratios between 1915–1917 and 1918–1919 and 2.5 percent of the variation in the change in their claims per \$1,000 of insurance in force.

This paper contributes to three literatures. The first examines whether geographic diversification reduces the risk of financial intermediaries (Demsetz and Strahan, 1997; Goetz et al., 2016; Carlson and Mitchener, 2006). I show that the value of geographic diversification depends on the correlation of an intermediary’s risks within a place, which was low for insured lives and is high for the local credit risks of banks, and that the expansion through which an intermediary becomes diversified can raise the cost of its underwriting. The second is the literature on the corporate diversification discount, which asks whether diversified firms are worth less because diversification makes the firm as a whole less efficient or because they hold segments that would be worth less in any firm (Lang and Stulz, 1994; Berger and Ofek, 1995; Campa and Kedia, 2002; Villalonga, 2004). The exact decomposition answers the analogous question for an intermediary market by market, and it can be applied to any firm observed by market, such as banks by branch or mortgage lenders by county. The finding that expansion raised claims throughout the firm also bears on the costs of transmitting soft information through a hierarchy (Stein, 2002; Liberti and Mian, 2009) and on the dilution of monitoring in a diversified lender (Winton, 1999). The third is the history of American finance, to which the paper adds a market-level account of how a major intermediary became national (North, 1954; Davis, 1965; Snowden, 1995; Murphy, 2010; Arthi et al., 2025; Janas,

2026). Research on insurers’ diversification has focused on diversification across lines of business rather than across places (Liebenberg and Sommer, 2008; Che and Liebenberg, 2017; Berry-Stölzle et al., 2012), and research on the fragility of life insurers has focused on nontraditional liabilities (Foley-Fisher et al., 2020; Koijen and Yogo, 2022).

The paper is organized as follows. Section 2 describes the organization of the industry and sets out a framework for the variance of an insurer’s claims. Section 3 describes the data and the measures. Section 4 documents the geographic expansion of the industry, Section 5 the cost of that expansion, and Section 6 its consequences for stability. Section 7 examines the 1918–1919 influenza pandemic, and Section 8 concludes.

2 Institutional Background and Conceptual Framework

2.1 The organization of life insurance and the mode of expansion

Life insurance was the largest nonbank conduit between American households and long-term credit markets in the decades around 1900 (Stalson, 1942). The ordinary policy combined insurance and saving in a single contract (Arthi et al., 2025): the insurer charged a level premium that exceeded the cost of current mortality in the early years of a policy, accumulated the excess as a policy reserve, and invested the reserve in mortgages, municipal bonds, and corporate securities, becoming a major supplier of long-term credit (Snowden, 1995; Davis, 1965).

Insurance was regulated by the states rather than the federal government throughout the period, and an insurer operated in a state only under a license from the state’s insurance department. A license typically required a deposit of securities or evidence of the insurer’s solvency under the state’s reserve standard, and states’ licensing requirements, reserve standards, and investment rules varied widely. Once licensed, an insurer sold policies through local agents, who solicited applications and collected premiums within their territories. Underwriting, reserving, and investment policy were nevertheless set at headquarters under a single charter. An application was solicited by a local agent and supported by the report of a local medical examiner, and the home office accepted or declined it on the basis of those reports. Over the period the selection of risks was progressively standardized across the industry. The Medical Information Bureau, founded by the companies’ medical directors in 1902, pooled their records of impaired applicants; the Medico-Actuarial Mortality Investigation, published in 1912–1914, measured on the pooled experience of the major companies the mortality associated with build, occupation, and medical impairments; and the numerical rating

system developed at the New York Life and published in 1919 scored each element of an applicant's risk against such standards (Association of Life Insurance Medical Directors and Actuarial Society of America, 1912–1914; Rogers and Hunter, 1919; Bouk, 2015).

An insurer therefore expanded geographically by entering states one at a time. It was chartered in one state, obtained licenses in others, appointed agents there, and wrote its first policies. Each new position opened at small scale, because the insurer began with no insurance in force in the state and its business grew only as quickly as its new agents sold policies, and it was initially favorably selected, because newly examined lives incur few claims in the first years after selection. Outside its home state an insurer lacked an established agency, a local reputation, and the local market share that would confer pricing power, so it wrote at a lower margin. Most entries remained small or were abandoned, and a few grew into large markets. The geographic breadth of an insurer at maturity is the cumulative outcome of these entries, so breadth and size are joint outcomes of a single process of expansion. Two insurers of the same size could therefore have very different geographic distributions of insurance in force, one concentrated in a few states and another spread across many, for reasons attributable to the history of their agency networks.

The 1905–1906 Armstrong investigation in New York, a response to scandals at the largest insurers, tightened oversight, disclosure, and investment rules, and because New York companies dominated the industry its reforms reshaped practice nationally (North, 1954). The investigation followed two decades in which the largest companies competed for new business through deferred-dividend (tontine) policies and costly agency forces (Keller, 1963; Ransom and Sutch, 1987; Stalson, 1942); the legislation ended new deferred-dividend business and limited the expenses and the volume of new business of the largest New York companies.

2.2 The variance of claims and the insurer's balance sheet

The stability of a life insurer depends on the predictability of its claims. An insurer sets its premiums and reserves on the basis of the mortality it expects among the lives it insures. Its policy liabilities, the reserves held against future claims, are valued under an assumed mortality table and interest rate, and its surplus, the excess of assets over liabilities, absorbs the difference between realized and expected claims. When claims are more variable, an insurer requires a larger surplus to maintain the same probability of solvency, and a given surplus provides less protection to policyholders. The variance of claims also affects the insurer's investment policy. An insurer whose claims are predictable can match the timing of its claim payments with premium income and the income from

its investments, and can therefore hold long-term, illiquid assets such as mortgages; an insurer whose claims are less predictable must hold more liquid assets or risk selling long-term assets at a loss to meet claims. Finally, for the participating policies that dominated the industry, variation in claims was transmitted to policyholders through the dividends paid from surplus.

The variance of claims plays the same role for an insurer that the variance of loan losses plays for a bank. In each case the intermediary's capital absorbs unexpected losses on its portfolio, so that a less variable portfolio supports the same claims on the intermediary with less capital (Diamond, 1984). The liquidity problem differs in form: a bank must meet withdrawals of demand deposits, which depositors can make at any time and may make simultaneously (Diamond and Dybvig, 1983), whereas an insurer must meet claims whose timing is determined by the mortality of the insured. For an insurer, therefore, the predictability of claims governs both the capital it requires and the liquidity it must hold. I measure stability through the variance of the loss ratio, claims relative to premiums, which normalizes claims by the size of the insurer's business, and through the insurer's survival and the behavior of its surplus.

2.3 Sources of claims variance: pooling across individuals and across locations

The variance of an insurer's claims depends on how the claims of the lives it insures are correlated, and the relevant correlation is between lives in the same place. Consider an insurer with n insured lives, of which $n_s = w_s n$ reside in place s , so that $\sum_s w_s = 1$. Let the claim on life j , who resides in place $s(j)$, be

$$x_j = \mu + u_{s(j)} + e_j,$$

where u_s is a component common to all lives in place s , with variance τ^2 , and e_j is a component specific to the individual, with variance $\sigma^2 - \tau^2$. The place components are independent across places, the individual components are independent across individuals, and the two are independent of each other. The variance of a single claim is therefore $\text{Var}(x_j) = \tau^2 + (\sigma^2 - \tau^2) = \sigma^2$. Two distinct lives j and k in the same place share only the place component, so $\text{Cov}(x_j, x_k) = \text{Var}(u_s) = \tau^2$, and the correlation between their claims is

$$\rho \equiv \text{Corr}(x_j, x_k) = \frac{\tau^2}{\sigma^2}, \quad s(j) = s(k), \quad j \neq k.$$

The claims of lives in different places are uncorrelated.

The insurer's average claim per life is $\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j$. Because the place component u_s enters

the sum once for each of the n_s lives in place s ,

$$\frac{1}{n} \sum_{j=1}^n u_{s(j)} = \sum_s \frac{n_s}{n} u_s = \sum_s w_s u_s,$$

so that

$$\bar{x} = \mu + \underbrace{\sum_s w_s u_s}_{\text{place component}} + \underbrace{\frac{1}{n} \sum_{j=1}^n e_j}_{\text{individual component}}.$$

The two components are independent, so the variance of $\bar{x}$ is the sum of their variances. Because the place components are independent across places, the covariances among them vanish, and the variance of the place component is $\sum_s w_s^2 \text{Var}(u_s) = \tau^2 \sum_s w_s^2$. The individual component is the mean of n independent terms, each with variance $\sigma^2 - \tau^2$, so its variance is $n(\sigma^2 - \tau^2)/n^2 = (\sigma^2 - \tau^2)/n$. Hence

$$\text{Var}(\bar{x}) = \frac{\sigma^2 - \tau^2}{n} + \tau^2 \sum_s w_s^2,$$

and, dividing by the variance of a single claim and substituting $\rho = \tau^2/\sigma^2$,

$$\frac{\text{Var}(\bar{x})}{\sigma^2} = \frac{1 - \rho}{n} + \rho \sum_s w_s^2. \quad (2.1)$$

The first term arises from the individual component of claims and falls with the number of lives, wherever they reside: a large portfolio of policies written in one state is as well pooled in this respect as a large portfolio written in thirty. The second term arises from the place component and falls only with breadth, $1/\sum_s w_s^2$, the effective number of places in which the insurer writes. The two limiting cases illustrate the distinction. If $\rho = 0$, the variance is $1/n$, the familiar result for independent risks, and the geographic distribution of the insured lives is irrelevant. If $\rho = 1$, the variance is $\sum_s w_s^2$ regardless of the number of lives, and only diversification across places reduces it.¹

The two terms of equation (2.1) correspond to quantities that can be measured for each insurer. A death claim is $x_j = FD_j$, where F is the amount insured and D_j indicates the death of life j , which occurs with a small probability q . In the absence of place components the variance of total claims is $nq(1 - q)F^2 \approx nqF^2$, which equals expected losses, nqF , times F . Dividing by

¹Equation (2.1) omits a component common to all places, such as a national epidemic. If each claim also contains a component v with variance κ^2 that is common to every life, v enters $\bar{x}$ with total weight $\sum_s w_s = 1$, and equation (2.1) acquires an additional term κ^2/σ^2 that neither the number of lives nor their geographic distribution reduces. Section 6 returns to this term.

the square of premium gives the variance of the loss ratio that arises from the number of claims, $N \approx \text{losses} \times F/\text{premium}^2$, the empirical counterpart of the first term, which Section 3.3 defines for each insurer. The counterpart of the second term is G , the variance of the insurer’s loss ratio implied by the premium-weighted mortality shocks of the states in which it writes, also defined there.

Equation (2.1) implies that, conditional on size, geographic breadth stabilizes an intermediary in proportion to ρ , the correlation of risks within a place. The correlation of risks across places, by contrast, determines the part of the variance that neither size nor breadth can remove. For a life insurer, local mortality shocks arise from epidemics confined to particular cities or states, such as typhoid and diphtheria outbreaks or the uneven severity of the 1918 influenza pandemic; from natural disasters, such as floods, hurricanes, earthquakes, and fires; and from mining and industrial accidents that could kill many workers in a single locality at once. Such events cause the claims of many policyholders in one place to occur together, and an insurer whose policies were concentrated in an affected state would have experienced a large increase in claims.

For a bank, ρ is plausibly larger. A bank’s borrowers in a market depend on the same local economy, so that a decline in local crop prices or employment raises defaults on many loans at once, and a bank’s depositors in a market may withdraw together. Goetz et al. (2016) find that geographic expansion reduces bank risk by reducing exposure to such idiosyncratic local risks, and particularly so when banks expand into economically dissimilar markets, which is the prediction of equation (2.1) when ρ is large. In the theory of delegated monitoring, diversification across loans whose returns are independent reduces the costs of monitoring on behalf of depositors (Diamond, 1984); what matters there is the number of independent risks, not their location, and equation (2.1) makes the distinction explicit. Section 6 measures both terms for each insurer.

2.4 Geographic expansion and the selection of risks

Expansion may also affect how well an insurer selects its risks. Underwriting was administered centrally but executed locally: the home office decided on applications solicited and examined by local agents and medical examiners. If screening depends on soft information, which “cannot be directly verified by anyone other than the agent who produces it” (Stein, 2002), the quality of selection depends on the attention that management devotes to the agents and examiners in each market. Management’s attention is limited, so an insurer operating in thirty states devotes less of it to each than an insurer operating in three, and it screens less effectively in each (Liberti and Mian,

2009; Agarwal and Hauswald, 2010; Winton, 1999). Expansion into new markets also demands attention, to licensing, to the appointment of agents, and to the supervision of new agencies, that would otherwise be devoted to the markets the insurer already served.

Limited managerial attention has three implications that distinguish it from a friction specific to the markets an insurer entered. First, it raises claims throughout the insurer’s business, including its headquarters state and the markets it served before expanding, rather than only in entered or distant markets. Second, it raises claims in an insurer’s incumbent markets when the insurer expands elsewhere. Third, its cost should decline as selection is standardized, because standardized procedures reduce the dependence of selection on the attention of management. Section 5 examines each implication.

3 Data and Measurement

3.1 The state-level panel

The principal source is the *Insurance Year Book*, published annually by the Spectator Company, which reports the premiums collected and losses paid by every legal-reserve life insurer in every state in which it operated, together with new insurance written, insurance in force, each insurer’s headquarters, and, for insurers that left the industry, the year and channel of exit. I transcribe these tables for 1888–1940 into a firm-by-state panel. Fraternal benefit societies are not covered. The panel contains 149,581 firm-state-years, or 10,952 firm-years of 818 insurers; ordinary business that passes the validity screen described below, 123,870 firm-state-years and 10,341 firm-years of 781 insurers, is the analysis sample.² Throughout the paper, a market is a state, and an insurer’s markets are the states in which it collected premium in a given year.

(Table I about here)

The data are highly skewed (Table I). The median insurer collected about \$574 thousand in premiums and the mean \$6.7 million, a difference attributable to a few large national insurers. The same insurers account for the skewness in geographic breadth: the median insurer wrote in 6 states but had an effective number of states, defined below, of only 2.15, so that even insurers licensed in many states concentrated their premiums in a few home markets. A typical insurer paid death

²Online Appendix Section A reconciles the panel with the analysis sample and documents the transcription, coverage, exits, the validity screen, and the validation against the New York statements.

claims of about \$7 per \$1,000 of insurance in force, and its insured lives died at about three fifths of the rate assumed in its valuation table.

3.2 The insurers’ own accounts and the New York statements

The *Year Book* also printed each company’s accounts, including its death claims, number of deaths, policies and insurance in force by type, medical fees, and, in the gain-and-loss exhibit, its actual and expected mortality; I transcribe them for 1886–1926 for every insurer and link them to the panel. These accounts cover the small insurers that the New York statements omit: the median firm-year carried \$12 million of insurance in force on about 7,200 policies. The annual statements filed with the New York Superintendent of Insurance add balance sheets, income statements, and policy exhibits for the roughly 170 insurers licensed in New York, of which 75 can be linked to the panel; the matched samples contain 41 to 75 insurers, depending on the exhibit, and consist mostly of large national companies.

3.3 Measuring breadth, size, entry, and the components of claims variance

I summarize each insurer’s geographic distribution with a single measure, *breadth*, built from the effective number of states $N_{\text{eff}} = 1/\sum_s w_s^2$, the inverse Herfindahl index of the shares w_s of an insurer’s premium collected in each state s . In every regression, breadth is the standardized log of N_{eff} , and I use N_{eff} itself only to describe the data. *Size* is the standardized log of total premium. Unless the notes to an exhibit state otherwise, both are standardized within the sample of each estimate, and within each period where periods are estimated separately; the standard deviation of log N_{eff} is 0.96 over 1888–1920 and 0.93 over 1888–1940, so the units are comparable. In the analysis sample N_{eff} averages 4.43 effective states, with a median of 2.15.³

The first year of the panel complicates the dating of an insurer’s founding and of its entry into each state. For an insurer already active in 1888 I observe neither its founding nor the dates at which it entered the states it already served. For the analyses of Section 4, I therefore treat any firm-state present in 1888 as pre-existing, date an insurer’s founding to the first new markets entered in 1889, and count entries from 1890. Breadth, size, and all firm-level outcomes are computed from all of an insurer’s markets, and the market-level regressions of Section 5 retain markets held in 1888 with an indicator. Of the 781 insurers in the analysis sample, 729 are first observed in 1890 or later.

³The results are robust to measuring breadth by the level rather than the log of N_{eff} and to bins of breadth (Online Appendix Section B).

An *entry* is a firm-state’s first observed year with positive premium, restricted to new business by two conditions: most of the insurance in force in the market must be newly written (new insurance written over insurance in force above one half), and the initial position must be small (premium no larger than the insurer’s median state premium that year). The conditions exclude blocks of policies assumed through reinsurance and markets in which an insurer was already large. A market’s *tenure* is the number of years since its entry, and the markets an insurer served before a given year are its *incumbent* markets in that year.⁴

To measure the two terms of equation (2.1), I decompose the variance of each insurer’s loss ratio over five-year windows of consecutive years. Let R denote the realized variance of the insurer’s loss ratio in the window. With n policies of average size F paying claims at a small rate, the number of claims is approximately Poisson, and the variance of the loss ratio attributable to randomness in that number is $N \approx \text{losses} \times F/\text{premium}^2$, which is inversely proportional to the number of lives insured, wherever they reside. N corresponds to the first term of equation (2.1). The residual, $R - N$, reflects fluctuations in mortality and in the composition of claims, and it contains the second term, the only component that diversification across states can reduce. To measure that component directly, I compute G , the variance the insurer’s loss ratio would have had from the mortality of its markets alone: for each state in which it wrote, the loss ratio of the other insurers in that state and year less the national loss ratio of the year, weighted by the insurer’s premium in the state.

Market cells whose figures are implausible for life insurance, 1.9 percent of firm-state-years and 0.9 percent of premium, are set aside before any firm total, share, or breadth measure is built, and the evidence indicates that they are reporting errors. Because the excluded cells are concentrated after 1920 and the attribution of the full-period estimates depends on their treatment, the principal estimates of Section 5 are those for 1888–1920.⁵

4 The Geographic Expansion of U.S. Life Insurers

This section documents how the industry grew, where its growth occurred, and the process through which individual insurers became national. The evidence establishes that geographic breadth was an outcome of growth, which motivates the analyses of Sections 5 and 6.

⁴Some insurers first observed after 1889 are older companies that the *Year Book* began to itemize late. The results of Section 4 are robust to restricting the sample to insurers covered continuously from their first appearance (Online Appendix Section A).

⁵Online Appendix Section A gives the evidence that the excluded cells are reporting errors; the main results are robust to alternative treatments of them (Online Appendix Section C).

4.1 The growth of the industry

(Figure I about here)

The number of active legal-reserve life insurers rose from about 50 in 1888 to about 350 around 1930, and premium income rose from about 0.7 percent of GDP in 1890 to a peak of 4.3 percent in 1935 (Figure I, Panels A and B). Entry accelerated after 1900, particularly in the Midwest and the South, where the industry had previously been served almost entirely by companies headquartered in the Northeast. Panel B also reports the assets of life insurers relative to those of commercial banks.

4.2 Sources of premium growth: incumbent and entered markets

(Table II about here)

The growth of industry premium can be attributed to four sources: the growth of insurers' incumbent markets, the markets they served at the start of a decade; markets entered during the decade by insurers already in operation; insurers that first appear during the decade; and markets abandoned during the decade. Table II reports this decomposition by decade, with premium converted to constant dollars and each source expressed in percent of industry premium at the start of the decade. Real industry premium grew by 137 percent in the 1890s, 70 percent in the 1900s, 17 percent in the 1910s, when the price level more than doubled, 181 percent in the 1920s, and 17 percent in the 1930s. In every decade most of this growth came from markets insurers already served at its start, which added 98, 57, 11, 171, and 18 percent of starting premium in the five decades. Because an entered position's premium grew more than threefold in its first decade (Section 4.4), much of this growth represents the expansion of positions entered in earlier decades. Markets entered during the decade added 32 percent of starting premium in the 1890s and between 2 and 7 percent in each later decade, insurers first appearing added between 3 and 13 percent, and abandoned markets subtracted between 2 and 6 percent. The premium of entered markets came predominantly from the South and the Midwest, which together accounted for 68 percent of it in the 1890s, 73 percent in the 1900s, and 83 percent in the 1910s (Panel B).

(Figure II about here)

Market entry was frequent and widespread. Figure II reports the number of entries by decade and by the census region of the entered state. Entries rose from 182 in the 1890s to 516 in the 1900s,

614 in the 1910s, and 958 in the 1920s, before falling to 394 in the 1930s. Relative to the number of insurers, entry was most frequent in the 1900s, at 46 entries per 100 firm-years, and least frequent in the 1930s, at 13. In every decade most entries were into the South and the Midwest, which received between 64 and 75 percent of them, whereas entries into the Northeast, where the industry originated, were few.

4.3 The geographic structure of the industry

(Table III about here)

The geographic structure of the industry reflects two concurrent processes: the expansion of established insurers into new states and the formation of new insurers, most of them local, in the states into which the established insurers were expanding. Table III reports the share of premium an insurer collected in its headquarters state, the share of each state’s premium written by insurers headquartered elsewhere, and the effective number of states. Weighted by premium, the headquarters-state share was between 0.20 and 0.22 in every year from 1890 to 1940, and the premium-weighted effective number of states was between 9 and 11 from 1900: the typical premium dollar was collected by an insurer that wrote most of its business outside its home state. The median insurer, by contrast, became more local. Its headquarters-state share rose from 0.28 in 1890 to 0.80 in 1910 and was about 0.60 thereafter, and its effective number of states fell from 4.1 to 1.4 before recovering to about 2, because the insurers formed after 1900 began in their home states. Insurers headquartered in other states wrote 0.94 of the premium of the average state in 1890; the formation of local insurers reduced this share to 0.87 by 1920, after which it recovered slightly. The typical premium dollar was collected by an insurer writing in about 40 states (Figure I, Panel D).

(Figure III about here)

Local markets became considerably more competitive without displacing the national leaders. The average number of insurers operating in a state rose from about 17 in the late 1880s to nearly 80 by 1940, and the share of a state’s premium written by its five largest insurers fell from about 80 to about 50 percent (Figure III, Panels A and B). Insurers in the top decile of national premium nevertheless collected 80 to 90 percent of it from the early 1910s through 1940 (Panel C), and they were largely the same insurers throughout.⁶

⁶Janas (2026) documents the industry’s market structure in detail. Online Appendix Section C reports the persistence of the leading insurers and the annual series behind Table III.

4.4 The process of expansion: the scale, claims, and prices of entered markets

(Table IV about here)

Within an insurer, premium migrated from the markets in which it was founded to those it entered as it aged (Table IV). Founding states account for 68 percent of premium at firm age 5 and 43 percent at age 20, so that by age 20 an insurer collected nearly three of every five premium dollars in markets it entered after its founding, and the earlier a market was entered, the larger its share.

(Table V about here)

Most insurers never attained a national scale. Table V reports, for insurers first appearing in each decade, the share that reached 3, 5, and 10 effective states within ten years and the median number of years required among insurers that reached each threshold. Of the 147 insurers first appearing in the 1900s, 36 percent reached 3 effective states within a decade and 4 percent reached 10; of the 182 first appearing in the 1910s, 29 and 6 percent; and of the 190 first appearing in the 1920s, 42 and 12 percent. The 31 insurers first appearing in the 1890s, which include older companies that the *Year Book* began to itemize late, expanded most widely: 64 percent reached 3 effective states within a decade and 20 percent reached 10. Among the insurers that reached three effective states, the median insurer did so within two to eight years of its first appearance, so the distinction between national and regional insurers was largely established early in an insurer's life.

The gradual migration of premium toward entered markets reflects the development of each entered position. Figure IV follows entered markets along two dimensions, their tenure and their distance from headquarters, and reports two outcomes, premium (Panels A and B) and the loss ratio (Panels C and D). Each comparison is among the markets of a single insurer in a given year, so it is not affected by differences across insurers in size, organization, or the quality of management. The panel titles state the reference group of each comparison, which differs between the two tenure profiles: Panel A follows a market relative to its own first two years, whereas Panel C compares entered markets with the insurer's founding states.

(Figure IV about here)

Entered positions opened at small scale and grew with tenure. Among the markets an insurer continued to serve, and relative to the rest of its business, a position's premium reached nearly three times its initial level within 5 years and about three and a half times after a decade

(Panel A). Distance from headquarters limited the scale of a position: an entered market more than 2,000 kilometers from headquarters carried about a third of the premium of one within 250 kilometers (Panel B).

New positions were also favorably selected, and the advantage dissipated as they grew. In its first two years a newly entered market had a loss ratio about 11 points below that of the insurer’s founding states, and the difference narrowed with tenure, to 2 points after a decade (Panel C). The two profiles are related. A new position is small, and a market with few insured lives is more likely by chance to record no claims in a given year. Controlling for the log of the market’s insurance in force, which within a firm-year is equivalent to controlling for the number of lives insured, reduces the initial difference by about two fifths, and the remaining difference still narrows with tenure. New positions were therefore favorably selected, but the growth of their insurance in force overstates the rate at which this advantage dissipated. Among entered markets of the same age, claims experience was similar across the first four distance bands, and only markets more than 2,000 kilometers from headquarters had lower loss ratios, by about 3 points (Panel D).⁷

(Table VI about here)

Entered markets were also written at lower prices. Table VI estimates, over the firm-state-year panel,

$$y_{ist} = \beta \ln d_{is} + \gamma \tau_{ist} + \alpha_{it} + \varepsilon_{ist}, \quad (4.1)$$

where y_{ist} is the market’s loss ratio, its premium per \$1,000 of insurance in force (premium yield), or one of the two margins of the loss ratio (the probability of observing no claims and the loss ratio conditional on at least one claim); d_{is} is the great-circle distance from the insurer’s headquarters city to the centroid of state s ; τ_{ist} is the market’s tenure; and α_{it} are firm-by-year fixed effects, which absorb all characteristics of the insurer in a given year, including its size and age. Standard errors are clustered by firm and by state-year.

Premium yield declines with distance from headquarters, by 0.74 per log-kilometer against a mean of 30.28, and the decline is robust to controls for tenure (0.58) and for the log of the market’s insurance in force (0.58). The loss ratio declines with distance as well: an insurer’s more distant markets have loss ratios 0.039 lower per log-kilometer against a mean of 0.276. Yield increases with an insurer’s share of a state’s premium, by 17.1 per unit of share, so that a one-standard-deviation increase in local share, about 0.04, raises yield by about 0.68, almost exactly offsetting the decline

⁷The pattern is robust to a Poisson model of losses with premium as exposure (Online Appendix Section C).

associated with a log-kilometer of distance (Panel B). The lower prices of distant markets therefore reflect the small local shares that an entrant held. When each market is weighted by its contribution to the insurer’s totals (Panel C), the loss-ratio and claims gradients fall by half or more while the yield gradient falls only by a fifth: the favorable claims experience of distant markets is concentrated in positions too small to affect an insurer’s aggregate outcomes, whereas the lower prices hold in markets of every size.⁸

(Table VII about here)

The cost to an insurer of the lower prices of entered markets can be measured directly. Within a firm-year and a state-year, a market the insurer entered after its first year earned 0.47 less per \$1,000 of insurance in force than its original markets (standard error 0.14), against a mean of 30.6, and the sign of the difference reverses once the insurer’s local share in the market is held fixed (Table VII, Panel A). Applied to the insurance in force of entered markets and discounted at 4 percent, the premium forgone was 1.5 percent of the premium those markets earned, with a 95 percent confidence band of 0.6 to 2.5 percent, and 1.3 percent before 1920. Relative to premium loadings of 10 to 20 percent, the pricing cost of new positions was economically small.

An entered market therefore began as a small, favorably selected position, written at a lower price because the entrant held a small share of the local market. As the position grew, its claims experience converged toward that of the insurer’s other markets, while its price remained lower for as long as the insurer’s local share remained small.

4.5 The distribution of returns to market entry

(Figure V about here)

The entries that succeeded were a minority that could not be identified in advance. Figure V follows each entry for ten years. An entry that the insurer ceased to serve has a payoff of zero; for the others, the payoff is the growth of the market’s premium relative to the growth of the insurer’s total premium, measured from the second year of operation, so that a value of one denotes a market that grew at the same rate as the rest of the insurer’s business.⁹ About a fifth of entries were abandoned

⁸Most of the unadjusted loss-ratio gradient operates through whether any claims occur, a margin that disappears once the market’s insurance in force is held fixed. The gradients are robust to measuring distance to each state’s largest city (Online Appendix Section C).

⁹The entry year is often a partial year of operation, so payoffs measured from it overstate later growth; the abandonment rate and the concentration of premium are robust to this choice. Online Appendix Section C reports these results and the regressions of an entry’s payoff on the information available at entry.

within the decade, the median surviving entry grew about twice as fast as the rest of the insurer, and the best tenth of entries produced about half of a cohort’s ten-year premium, 48 percent even when entries are ranked within terciles of the size of the entered state.

The returns to entry were difficult to predict. A regression of an entry’s log payoff on the information available at entry (distance from headquarters, the insurer’s size and age, the market’s size, growth, and number of insurers, and the year) has an out-of-sample R^2 of 0.05. Adding the first year’s premium and early loss ratio raises it to about a third, and insurers acted on that information: a market whose premium grew a standard deviation more slowly than the insurer’s in its first two years was 2.1 percentage points more likely to be abandoned within the decade. The process resembles the sequential entry of exporters, who enter foreign markets at small scale and expand conditional on early success (Albornoz et al., 2012; Ruhl and Willis, 2017), with the difference that each position an insurer opened created policy liabilities extending over decades.

4.6 Breadth and size

(Figure VI about here)

Because insurers grew by entering markets, the largest insurers were also the broadest. Figure VI plots N_{eff} against the log of total premium. The effective number of states rises steeply and convexly with size, from about 1 among the smallest insurers to about 11 among the largest, and size deciles explain about 57 percent of its variation.

The evidence of this section raises two questions. The first concerns the cost of expansion. Entered markets were written at lower prices and, in their first years, had lower loss ratios than an insurer’s founding markets, so the composition of a broad insurer’s business differed from that of a narrow insurer. Whether expansion also changed the insurer itself, and in particular the claims it paid in all of its markets, cannot be determined from comparisons within an insurer’s markets, because any change common to all of an insurer’s markets is absorbed by the firm-by-year effects. Section 5 separates the firm-wide component of an insurer’s outcomes from the composition of its markets. The second question concerns stability. Because breadth and size moved together, the cross-section of insurers cannot reveal which of the two made the largest insurers stable. Section 6 separates them.

5 The Cost of Geographic Expansion

An insurer became broad only by entering markets, so a cost of breadth is a cost of geographic expansion. This section shows that insurers that expanded geographically paid more in claims per dollar of insurance in force than insurers of the same size, and that the difference was a property of the firm rather than of the markets it entered. Section 5.1 sets out an exact decomposition of an insurer’s outcomes into a firm-wide level and the composition of its markets. Section 5.2 shows that before 1920 the claims differential lay in the firm-wide level, held among insurers writing in the same market, and appeared in the death claims and deaths reported in the insurers’ own accounts. Section 5.3 examines whether claims in an insurer’s incumbent markets changed when it expanded elsewhere, and Section 5.4 that its incidence and timing are consistent with limited managerial attention to the selection of risks.

5.1 An exact decomposition of firm outcomes

Every outcome an insurer reports is an aggregate of the markets in which it writes, so a broad insurer’s loss ratio may differ from a narrow insurer’s because the two operate in different markets, because they hold different positions within those markets, or because they conduct business differently wherever they write. The first two are properties of the composition of the insurer’s business; the third is a property of the firm.

For insurer i writing in state s in year t , the market’s loss ratio can be written as

$$lr_{ist} = \underbrace{\phi_{it}}_{\text{firm-wide level}} + \underbrace{\alpha_{st}}_{\text{market}} + \underbrace{\delta(\tau_{ist}, d_{is})}_{\text{position within the insurer's markets}} + \varepsilon_{ist}, \quad (5.1)$$

where ϕ_{it} is common to every market the insurer serves in that year and captures all that is common to its business: the quality of its selection of risks and its pricing, but also the age and policy mix of the lives it insures; α_{st} is common to every insurer writing in state s in year t and captures local mortality, demand, regulation, and competition; and δ describes how a market differs from the insurer’s other markets as a function of its tenure τ_{ist} and its distance d_{is} from headquarters, the gradients documented in Section 4.4. An insurer’s aggregate loss ratio is the premium-weighted average of its markets, $LR_{it} = \sum_s w_{ist} lr_{ist}$ with w_{ist} the share of the insurer’s premium collected in state s . When equation (5.1) is estimated by least squares with each market weighted by w_{ist} , the firm-by-year effects force $\sum_s w_{ist} \hat{\varepsilon}_{ist} = 0$ within every firm-year, and the aggregate decomposes

exactly:

$$LR_{it} = \underbrace{\hat{\phi}_{it}}_{\text{level}} + \underbrace{\sum_s w_{ist} \hat{\alpha}_{st}}_{\text{markets}} + \underbrace{\sum_s w_{ist} \hat{\delta}(\tau_{ist}, d_{is})}_{\text{position}}. \quad (5.2)$$

Any firm-level outcome that is a weighted average of the outcomes of its markets, such as claims or premium per \$1,000 of insurance in force, decomposes in the same way. The decomposition is exact only with the weights that define the aggregate, premium shares for the loss ratio and shares of insurance in force for claims and premium per \$1,000; with equal weights across markets the residual would enter the aggregate as a fourth term. The markets and position terms together form the *composition* of the insurer's business.

Breadth enters equation (5.2) through all three terms. Composition varies with breadth mechanically, since breadth is itself a function of the weights w_{ist} : an insurer cannot become broader without shifting weight from the markets in which it was founded toward those it entered, which differ in tenure, in distance, and in the markets themselves. Through the level term, breadth may covary with the quality of an insurer's underwriting across all of its markets, in which case it alters the aggregate without any change in composition. A regression of a firm-level outcome on breadth identifies the sum of these terms, and the terms need not have the same sign. Section 4.4 established that entered markets had lower loss ratios than founding markets, a composition effect; the decomposition determines whether a firm-wide difference offsets or reinforces it.

The question is the same one that divides the literature on the corporate diversification discount: whether diversified firms are worth less because diversification makes the firm as a whole less efficient or because they hold segments that would be worth less in any firm (Lang and Stulz, 1994; Berger and Ofek, 1995). Berger and Ofek (1995) compare a diversified firm's value with the sum of the imputed stand-alone values of its segments, and later work asks how much of the resulting discount reflects which firms choose to diversify (Campa and Kedia, 2002; Villalonga, 2004). The markets term of equation (5.2) plays the role of the imputed stand-alone value: it is what the insurer's markets yield for the insurers writing in them, and the level term is the insurer's departure from it. Because the level and the composition of an insurer's business are observed market by market, the separation follows from an accounting identity rather than from a model of selection, and the design can be applied to any intermediary observed by market, such as bank deposits by branch or mortgage lending by county.

Identifying the terms requires observing an insurer market by market. A design with firm-by-year fixed effects absorbs ϕ_{it} , so the remaining variation is across the markets a single insurer serves

in a year and identifies δ . A design with state-by-year fixed effects absorbs α_{st} , so the remaining variation is across the insurers writing in the same market and year and identifies the covariance of ϕ_{it} with breadth. Estimating equation (5.1) with both sets of effects and aggregating with equation (5.2) yields all three terms simultaneously, on one sample and in the units of the aggregate, so that their breadth coefficients sum to the aggregate coefficient. The design is a weighted version of the two-way fixed-effects decomposition used to separate worker and firm effects in wages (Abowd et al., 1999), with firm-years in the place of workers and state-years in the place of employers, and it shares that design’s known limitation. A firm-year effect is estimated from the markets the insurer serves in that year, so it is imprecise when they are few, and its sampling error can be correlated with the other estimated effects (Andrews et al., 2008). Because narrow insurers are small, their firm-year effects are also the least precise; Section 5.2 reports the checks this requires.

5.2 The firm-wide claims differential

(Table VIII about here)

Table VIII reports the preferred specification and its exact decomposition. The outcomes are firm-year loss ratios, claims per \$1,000 of insurance in force, and premium per \$1,000 of insurance in force, in levels, because the decomposition is exact only in levels. Each coefficient is the effect of a standard deviation of breadth, conditional on a cubic in size, in percent of the outcome’s mean, with region-by-year fixed effects and standard errors clustered by firm.

Panel A reports 1888–1920. A standard deviation of breadth raised the loss ratio by 22.6 percent of its mean, raised claims by 18.7 percent, and lowered premium by 5.6 percent. The claims differential is attributable to the level term. Conditional on the markets it served and its positions within them, a broad insurer’s claims per dollar of insurance in force exceeded those of a narrow insurer of the same size by 21.8 percent; its markets and positions were slightly more favorable, so the composition of its business neither generated nor concealed the difference. After 1920 the claims differential is small (Panel B), 3.8 percent with a standard error of 3.6. In dollars, a standard deviation of breadth raised claims by 1.49 per \$1,000 of insurance in force before 1920, or 4.6 percent of premium; the claims attributable to breadth beyond what size predicts were 1.7 percent of the industry’s losses, material for the insurers that bore them but small for the industry.

The firm-level pricing differential is less robust. About two thirds of it is composition when local-share quintiles enter the position term, because a broad insurer held small shares in many states, and with size measured by a cubic in insurance in force, which is closer to the number of

lives an insurer pooled, it disappears: -0.6 percent (standard error 1.2) before 1920, against a claims differential of 20.8 percent (4.5). The pricing cost of expansion therefore rests on the comparisons within an insurer’s markets in Section 4.4, which do not depend on how size is measured.

The pre-1920 claims differential is robust to the checks the data allow. The division at 1920 was chosen after the estimates by five-year period (Figure VII) were in hand, and those estimates, rather than the division, describe when the differential was present: from 1888 to about 1915, declining thereafter. The estimates weight each firm-year equally. Weighted by lives or by insurance in force the claims differential is larger, but these estimates rest on the Spectator’s losses of the largest insurers, which exceed their own death claims in proportion to breadth (Online Appendix Section B), and I treat them as upper bounds. The differential is concentrated in the upper half of the size distribution.¹⁰

(Table IX about here)

The differential holds among insurers of the same size writing in the same state and year. Over the firm-state-year panel I estimate

$$y_{ist} = \theta \text{Breadth}_{it} + f(\text{Size}_{it}) + \lambda m_{ist} + \gamma \tau_{ist} + \alpha_{st} + \varepsilon_{ist}, \quad (5.3)$$

where breadth and size are measured at the firm-year level, $f(\cdot)$ is a cubic, m_{ist} is the insurer’s share of the state’s premium, τ_{ist} is its tenure in the state, and α_{st} absorbs all factors common to a market in a year; standard errors are clustered by firm and by state-year. Because θ weights markets rather than firm-years, it measures how a broad insurer’s experience in a given market differs from that of a narrow insurer in the same market. Because it does not require estimating a separate effect for each firm-year, it is also not subject to the imprecision of the firm-year effects of equation (5.1) for insurers writing in few states.

Table IX reports θ by period, in percent of the outcome’s mean, with each market weighted equally. Panel A includes only the size cubic, and Panel B adds the local share, tenure, and an indicator for markets held in 1888. With size controls only, a standard deviation of breadth raised claims per \$1,000 of insurance in force by 26.7 percent of the mean in 1888–1905 and 17.4 percent in 1906–1920, and by 2.7 percent, not statistically significant, in 1921–1940; with the local share and

¹⁰The claims differential is robust to controls for firm age, tenure, incumbency in 1888, and the growth path; to logarithms, alternative trims, and a Poisson model of claim counts; to excluding single-state firm-years, firm-years near exit, short-lived insurers, and industrial insurers; to weighting by the precision of the level term; to alternative denominators; and to clustering by headquarters state (Online Appendix Section B).

tenure held fixed, the estimates are 24.1, 11.5, and 3.6 percent. The loss ratio rose by 28.7 and 18.7 percent of its mean in the first two periods, whereas the differences in premium per \$1,000, -3.2 and -1.1 percent, are small and not statistically significant, so that before 1920 the within-market differential lay in claims rather than in prices. The within-market claims estimates for 1888–1920 are of the same magnitude as the level term of Table VIII, 21.8 percent, and they disappear after 1920, as the level term does. The difference is not accounted for by firm age or incumbency in 1888, by the five largest insurers, or by headquarters-state-by-year effects, which absorb the regulatory and agency environment of the insurer’s domicile and reduce the pooled loss-ratio and claims coefficients by about two fifths.¹¹

The differential is a difference in mortality rather than in reporting or in the type of policies sold. The state-table losses can include payments other than death claims, such as matured endowments, but the insurers’ own company accounts, which the Spectator printed for every insurer, show the same pattern: a standard deviation of breadth raised insurers’ own death claims per \$1,000 of insurance in force by 15.4, 17.5, and 13.8 percent of the mean in 1888–1905, 1906–1920, and 1921–1926, and deaths per 1,000 policies in force by 16 to 17 percent in every period. Among the insurers licensed in New York, whose annual statements report death claims directly, the difference before 1906 is 21.0 percent of the mean (standard error 8.0). Neither the policy mix, measured by the endowment share and the average policy size, nor participating business, including the deferred-dividend policies of the largest companies before 1906, accounts for the differential, although differences in the duration of the policies in force cannot be fully excluded. The evidence describes the typical insurer rather than the typical premium dollar: weighted by premium, the difference in own death claims is smaller and not statistically significant, so estimates of the claims differential that weight firm-years by their size are upper bounds.

5.3 Claims in incumbent markets around expansion episodes

The comparisons of Section 5.2 are across insurers. If expansion raised an insurer’s claims, claims in the markets it already served should have risen when it expanded elsewhere. I examine this implication with an event study of insurers’ incumbent markets around episodes of expansion. An *expansion episode* is the first year after an insurer’s first appearance in which it enters three or more states. For each episode year c , the treated insurers are those whose first episode occurs in c , and

¹¹The results are robust to weighting markets by their insurance in force (Online Appendix Section C), to controls for the endowment share and the average policy size, and to wild-cluster bootstrap inference (Online Appendix Section B).

the control insurers are those that have no episode or whose first episode occurs after $c + 6$; all are active in $c - 1$ in at least two states. The sample consists of these insurers' incumbent markets, the states in which they wrote before c , excluding each market's first year, over the years $c - 5$ to $c + 6$. Stacking the episode years, I estimate

$$y_{ist}^c = \sum_{k \neq -1} \beta_k \mathbb{1}[t - c = k] D_i^c + f(\text{Size}_{it}) + \lambda m_{ist} + g(\tau_{ist}) + \alpha_{is}^c + \alpha_{st}^c + \varepsilon_{ist}^c, \quad (5.4)$$

where D_i^c indicates the treated insurers of stack c , y_{ist}^c is claims per \$1,000 of insurance in force, $g(\cdot)$ is a set of tenure indicators, and α_{is}^c and α_{st}^c are market-by-stack and state-by-year-by-stack fixed effects. The coefficients β_k therefore compare, within the same state and year, the incumbent markets of an insurer that has just expanded with the markets of insurers that have not yet done so, relative to the year before the episode. Because the controls are insurers that have not yet expanded, the comparison avoids the bias of two-way fixed-effects estimators with staggered treatment. Standard errors are clustered by firm.

(Figure VIII about here)

Figure VIII reports the estimates for 97 treated insurer-episodes, and Online Appendix Section C reports the coefficients and the estimates for subsamples. Claims in an insurer's incumbent markets did not rise after an expansion episode. Pooled over years 0 to 6 after the episode, the coefficient is -0.17 per \$1,000 of insurance in force (standard error 0.37), against a mean of 6.60 before the episode, and the 95 percent confidence interval excludes an increase larger than 0.55 per \$1,000, about 8 percent of the mean and about a third of the firm-wide differential associated with a standard deviation of breadth before 1920 (Section 5.2). The estimate is similar for episodes in which the insurer entered three or four states and five or more (-0.18 and -0.14), with a control for the ratio of new insurance written to insurance in force (0.06), and for episodes after 1906 (0.19); for the few episodes in 1906 or earlier claims fell (-3.20 , standard error 1.51). In the years before the episode, claims in the incumbent markets of insurers about to expand were slightly lower than in the comparison markets (-0.63 , standard error 0.30, pooled over years -5 to -2), although the four leads are not jointly significant ($p = 0.12$).

The firm-wide differential of Section 5.2 therefore did not arise as a discrete deterioration in an insurer's existing markets at the time it expanded. The event study does not support the second implication of limited managerial attention set out in Section 2.4. It is consistent with a cost associated with the breadth an insurer had attained, which accumulated as its markets multiplied,

but it is also consistent with differences between the insurers that became broad and those that remained narrow that predate their expansion. The cost of geographic expansion documented in this section is therefore a difference between insurers that had expanded and insurers of the same size that had not, rather than an estimate of the effect of a given episode of entry.

5.4 Incidence and timing: limited managerial attention

(Figure VII about here)

The claims differential declined from about 1915 and had largely disappeared by the late 1920s (Figure VII). Estimated over five-year periods on the full-period sample, the breadth coefficient on the claims level term was between 19 and 31 percent of the mean in every period from 1888 to 1915 except 1896–1900, fell to 12.5 percent in 1916–1920 and 9 percent in 1921–1925, and was close to zero from 1926. The decline coincides with the standardization of the selection of risks across the industry described in Section 2.1, which reduced the dependence of selection on the attention of each insurer’s management.

The industry series combines two groups of insurers whose differentials did not coincide in time. Before 1906 it is attributable to the insurers licensed in New York: a standard deviation of breadth raised their claims level term by 35.0 percent of the mean (standard error 8.5), against 2.5 percent (10.7) among the other insurers, which were then few; in 1906–1920 the estimates are 14.7 (14.9) and 19.0 (5.3). The industry series does not fall at 1906, when the Armstrong legislation curtailed the new business of the largest New York companies, because the differential of the New York-licensed insurers remained high through 1906–1910 and was lower and imprecisely estimated thereafter, whereas among the other insurers, which entered the panel in large numbers after 1900, it fell from 30.3 percent in 1906–1910 to 20.2 and 9.0 percent in the next two periods. Their own death claims show the same division: in 1906–1920 the difference is 17.3 percent (4.7) among the other insurers and 1.7 percent (13.5) among those licensed in New York.

(Table X about here)

Within a broad insurer’s business the differential was present in every market and did not rise with distance from headquarters (Table X). The table re-estimates equation (5.3) for claims per \$1,000 by period, interacting breadth with indicators for the insurer’s headquarters state, for its largest market in the previous year when that market is not the headquarters state, and for the middle and farthest thirds of the other markets by distance from headquarters, so that the breadth

coefficient is the differential in the nearest third of the other markets. Before 1906 that differential is 3.65 per \$1,000 against a mean of 10.1. It is smaller by similar amounts at greater distances, in the headquarters state, and in the largest other market, so that a differential of about 2 per \$1,000 remains in every part of the insurer's business, and the decline with distance reflects the location of a broad insurer's nearest markets: most broad insurers were headquartered in the Northeast, the differential was larger in Northeastern markets, and with that difference allowed for (Panel C) the middle and farthest thirds no longer differ from the nearest. In 1906–1920 the differential is 1.30 per \$1,000 and does not vary with distance; it is smallest in the insurer's largest other market, although part of that difference is mechanical, since an insurer's own claims are a large part of the state-year mean in its largest market. After 1920 it is small. The differential does not vary with the insurer's tenure in a market, and it is smaller where the insurer's local share is larger only before 1906 (Panel B).¹²

The incidence and timing of the differential correspond to two of the three implications of limited managerial attention set out in Section 2.4. The differential was present throughout a broad insurer's business, including its headquarters state and the markets in which its business was concentrated, rather than only in entered or distant markets, and it declined, in markets at every distance, in the decade in which the selection of risks was standardized across the industry. Claims in an insurer's incumbent markets did not, however, rise when it expanded (Section 5.3). The favorable experience of new positions (Section 4.4) is consistent with this interpretation: that comparison is within an insurer's markets, where a new market contains recently examined lives whose claims have not yet accrued, whereas limited attention raises mortality in all of a broad insurer's markets relative to a narrow insurer's.

6 Geographic Expansion and Stability: Pooling Lives, Not Places

Broad insurers were more stable than narrow ones: their loss ratios were less volatile and their survival rates higher. By equation (2.1), conditional on size, breadth can add to stability only by diversifying risk that is correlated within places, and its contribution depends on how much of an insurer's risk is of that kind. This section shows that little of it was: insurers' stability came from the number of lives they pooled, not from the number of places in which those lives resided.

¹²The results are robust to measuring an insurer's claims against those of the other insurers in the same state and year and to defining the largest market by the current year's premium (Online Appendix Section C).

6.1 Stability conditional on size

(Table [XI](#) about here)

Conditional on size, breadth predicts neither volatility nor survival. The association between breadth and ten-year survival falls from 7.7 percentage points to 0.1 when the size cubic is included, while a standard deviation of size raises survival by 15.2 points (Table [XI](#), Panel A); the association between breadth and the log variance of the loss ratio, -0.59 unconditionally, becomes small and statistically indistinguishable from zero (Table [XII](#), Panel A). The unconditional survival advantage of broad insurers is not itself evidence of greater stability, because insurers left the industry by consolidation far more often than by failure. Of the 386 insurers with a recorded exit, 89 percent ceded their policies to another insurer through reinsurance, merger, or acquisition, 9 percent failed, and 2 percent retired (Table [XI](#), Panel B; Figure [I](#), Panel C), and the survival advantage of broad insurers consists of a lower probability of absorption by another insurer, not of failure. Nor did breadth contribute to stability through the scale it generated: holding earlier size fixed, earlier breadth predicts only a modest increase in later size, and holding later size fixed it predicts lower survival ten years later, by 4.2 percentage points (standard error 2.1), through more frequent consolidation with other insurers.¹³

For a financial intermediary the relevant risk is default on its liabilities, and insurers could fail through the asset side of the balance sheet as well as through their claims (Rose, 2017). In the annual statements of the insurers licensed in New York, broad insurers held less surplus relative to liabilities, and were less likely to experience a decline in surplus or to reduce policyholder dividends, but conditional on size, breadth predicts neither the level nor the volatility of surplus, whereas size predicts a less volatile surplus ratio (Table [XI](#), Panel C). The statements exclude only a large effect: with 80 percent power, a standard deviation of breadth would have to change the probability of a decline in surplus by 7 percentage points, against a mean of 32, to be detected, and for the surplus ratio and dividend reductions the minimum detectable effects are too large for the statements to be informative. The remainder of this section therefore concerns underwriting risk, the variability of claims relative to premiums. Insurers' assets, farm and urban mortgages and bonds concentrated in particular regions, were exposed to geography in ways their liabilities were not (Davis, 1965), and whether breadth on the asset side stabilized insurers is a question the state tables cannot answer.

¹³Online Appendix Section [C](#) reports exits by channel and the role of size in the relation between earlier breadth and later survival.

6.2 Pooling across lives and across places

(Table XII about here)

Size reduced both the claim-count component of an insurer's loss-ratio variance and the residual that geographic diversification could remove, whereas breadth, conditional on size, reduced neither. Table XII regresses the components defined in Section 3.3, N and $R - N$, measured over five-year windows, on breadth and size.¹⁴ In the specification with the size cubic (Panel A), a standard deviation of size lowers the log claim-count component by 1.98 and the log residual by 0.97, whereas the coefficient on breadth for the residual is 0.055, with a standard error of 0.068. Breadth is associated with a slightly higher claim-count component, but not through geography: at a given premium N rises with the loss ratio, and the entire coefficient on breadth for $\ln N$, 0.14, is accounted for by its coefficient on the log loss ratio, the claims differential documented in Section 5. The share of variance attributable to claim counts falls from 0.63 for small, narrow insurers to 0.08 for large, broad ones (Panel B) and, conditional on size, does not vary with breadth. The stability of large insurers reflected pooling across many lives.

The absence of an effect on the residual is partly a consequence of the level of the loss ratio. Breadth raised the window's mean loss ratio $\bar{L}$ by 0.144 log points per standard deviation (standard error 0.029), the claims differential of Section 5, and with proportional mortality shocks the variance of a loss ratio scales with the square of its level; across windows the elasticity of $R - N$ to $\bar{L}$, holding size fixed, is 1.70 (0.11). Measured relative to its level, as $\ln(R - N) - 2 \ln \bar{L}$, the residual variance falls with breadth, by 0.22 log points per standard deviation (0.06), and by 0.39 (0.09) in windows that start before 1917; controlling for $\ln \bar{L}$ instead of imposing the square gives -0.19 (0.06). Size lowers the relative residual as well, by 1.05 (0.09). Broad insurers' loss ratios were higher and, relative to that higher level, less dispersed; in absolute terms the two offset (Table XII, Panels C and D).

Geographic diversification accounts for little of either. Conditional on size, G , the variance attributable to the mortality of an insurer's markets, falls steeply with breadth, by 0.72 log points per standard deviation (standard error 0.06), but it is a small part of the variance at stake: in the median window G is 11 percent of $R - N$ (29 percent at the mean), and 5 percent for the largest third of insurers. If breadth had removed geographic variance from $R - N$ in proportion to its effect on G , a standard deviation of breadth would have lowered $\ln(R - N)$ by 0.06 at the median share

¹⁴The results are robust to retaining the windows in which $R - N$ is negative, using an inverse hyperbolic sine transformation, and to alternative measures of F and N (Online Appendix Section C).

and by 0.16 at the mean share; the estimate of 0.055 rejects both benchmarks at the 10 percent level (one-sided $p = 0.051$ and $p = 0.001$). Measured in proportion to the level of losses, the geographic share is smaller still, 4 percent in the median window and 16 percent at the mean, and implies reductions of 0.02 and 0.08 in the relative residual; the relative estimate of -0.22 exceeds both ($p = 0.001$ and $p = 0.014$). The lower relative dispersion of broad insurers' loss ratios is therefore larger than the removal of state-level mortality shocks could produce, and its source lies elsewhere, plausibly in claim-count variance that the measure of N , built for most firm-years from a common average policy size, leaves in the residual and whose size relative to the level falls as the level rises. Breadth did not make insurers of a given size more stable in absolute terms, and the geographic risk it could diversify was small relative to the variance arising from the number of claims.¹⁵

The same estimates indicate how large local mortality risk would have had to be for breadth to matter. With G falling by 0.72 log points per standard deviation of breadth, a standard deviation of breadth lowers the residual variance by the fraction $s(1 - e^{-0.72}) \approx 0.51s$, where s is the share of G in $R - N$. For breadth to lower the residual variance by a fifth, s would have to be about 0.39, more than three times its value of 0.11 in the median window. In the terms of equation (2.1), the within-state correlation of insured lives' claims was too low for geographic diversification to matter. Banks differ in this respect: a bank's borrowers in one market are exposed to the same local economy, and geographic expansion reduces bank risk most when banks enter economically dissimilar markets (Goetz et al., 2016).

6.3 Common shocks across an insurer's markets

(Figure IX about here)

The small contribution of breadth does not imply that an insurer's markets were uncorrelated. For each insurer writing in at least two states and each five-year window, let $B = \sum_s \bar{w}_{is}^2 \text{Var}(lr_{ist})$ denote the variance its aggregate loss ratio would have if its state loss ratios were uncorrelated. Then $1 - B/R$ is the share of its variance attributable to covariance across its markets. The median share is 0.09 in the lowest breadth quartile and rises to 0.24, 0.30, and 0.44 in the higher quartiles, with firm-block bootstrap standard errors of 0.02 to 0.04 (Figure IX, Panel B). Part of this increase is mechanical: diversification reduces the own-market variance terms, so the remaining variance of a broad insurer consists increasingly of the covariance component. State loss ratios covary because epidemics, the business cycle, and regional distress affect many markets simultaneously.

¹⁵Table XII, Panel E, and Online Appendix Section C.

The average pairwise correlation of annual state loss ratios is 0.23; year effects account for 14 percent of the variance of state-year loss ratios, and region-by-year effects for 42 percent, so the common component is largely regional. This common component is the third term omitted from equation (2.1), which neither size nor breadth reduces, and it limits what geographic diversification could have achieved even had local risk been larger.¹⁶

7 The 1918–1919 Influenza Pandemic

The estimates of Section 6 rest on the variation in mortality across states in ordinary years. The influenza pandemic of 1918–1919 provides a test of their interpretation under the largest geographically uneven mortality shock of the period. If insured lives’ claims were weakly correlated within states, a shock that raised mortality far more in some states than in others should have raised claims where it was severe, but differences in insurers’ geographic exposure should account for little of the variation in their claims, and breadth should have provided little protection.

7.1 The shock

The pandemic’s principal wave reached the United States in the autumn of 1918 and was the largest mortality shock of the period. Its severity varied substantially across states (Online Appendix Section C), and this variation has been used to study the pandemic’s economic consequences (Brainerd and Siegler, 2003). I measure the shock in each state by its excess deaths per 1,000 inhabitants in 1918, constructed from the mortality statistics of the death registration area (U.S. Bureau of the Census, 1920) and therefore available for the states in that area. Across the 25 states and the District of Columbia with data, excess mortality ranged from 1.6 per 1,000 in Wisconsin to 3.7 in Pennsylvania, Montana, and Colorado. Because the shock was unanticipated and the geographic distribution of insurers’ business was determined before it, the variation in insurers’ exposure is plausibly unrelated to other determinants of their claims.

7.2 Design

The analysis proceeds in two steps. The first estimates the effect of excess mortality on claims at the market level. For each insurer and state I compute the change in the loss ratio, and in claims per \$1,000 of insurance in force, between 1915–1917 and 1918–1919, and regress it on the state’s excess

¹⁶Online Appendix Section C reports the medians and their bootstrap standard errors. A pairwise correlation of this size, comparable to that of U.S. equities (Campbell et al., 2001), implies that common variation accounts for most of the variance of an insurer that writes evenly across forty markets.

mortality with insurer fixed effects, weighting markets by their 1915–1917 premium and clustering standard errors by state. The insurer fixed effects absorb any change common to all of an insurer’s markets, so the coefficient is identified from differences in excess mortality across the markets of a single insurer.

The second step examines the role of breadth. For each insurer, exposure is the premium-weighted average of excess mortality across the states in which it wrote, $E_i = \sum_s w_{is}^{1915-17} EM_s$, computed over states with mortality data, for insurers with at least half of their 1915–1917 premium in such states. Breadth affects the dispersion of exposure across insurers rather than its mean: a broad insurer’s exposure is close to the industry average, whereas a narrow insurer’s exposure depends on the few states in which it wrote. I first measure how much breadth reduced the dispersion of exposure. I then regress the change in each insurer’s loss ratio between 1915–1917 and 1918–1919 on its exposure, with a cubic in size and headquarters-region fixed effects. The product of the exposure coefficient and the difference in the standard deviation of exposure between narrow and broad insurers measures the protection that breadth provided against the shock. Three checks address the threats to this design. Exposure computed from 1912–1914 premium weights excludes any adjustment of insurers’ business in anticipation of wartime conditions; excluding the five states with the highest excess mortality ensures that the estimates are not determined by a few severely affected states; and a placebo regression of the change in the loss ratio between 1912–1914 and 1915–1917 on exposure tests whether insurers with high exposure already had rising loss ratios before the pandemic.

7.3 Results

(Table [XIII](#) about here)

The pandemic raised claims where it was severe. Within an insurer, a market’s loss ratio rose by 0.018 (standard error 0.010) for each additional excess death per 1,000 inhabitants, and claims per \$1,000 of insurance in force by 0.42 (0.38) (Table [XIII](#), Panel B). Across states, the aggregate loss ratio of all insurers rose by 0.030 (0.015) per excess death per 1,000. The pandemic was therefore visible in claims, but the market-level estimates are imprecise, because excess mortality varied only across 25 states and the District of Columbia, and a market’s loss ratio varies substantially from year to year.

Breadth reduced the dispersion of insurers’ exposure, but the dispersion it removed was small. Exposure converges toward the industry mean as breadth rises (Online Appendix Section [C](#));

its standard deviation is 0.71 excess deaths per 1,000 among the narrowest third of insurers and 0.28 among the broadest third, and a standard deviation of breadth reduces an insurer’s absolute deviation from the industry mean by 0.17 (standard error 0.04; Panel A). At the insurer-level coefficients reported below, this reduction in dispersion lowered the standard deviation of insurers’ loss-ratio changes by 0.005, against an actual standard deviation of 0.125, and the standard deviation of the change in their claims per \$1,000 by 0.90, against an actual 4.00 (Panel C). Differences in exposure account for 0.4 percent of the variation in the change in insurers’ loss ratios and 2.5 percent of the variation in the change in their claims per \$1,000.

At the insurer level, the coefficient of the change in the loss ratio on exposure is 0.011 with a standard error of 0.037 (Panel B). The estimate is consistent with the market-level coefficient but imprecise: its minimum detectable effect is 0.10, more than five times the market-level coefficient, because the variation in exposure across insurers is small relative to the variation in their loss ratios from other sources. The estimates with 1912–1914 weights and without the five most affected states are similar, and the placebo coefficient for 1912–1914 to 1915–1917 is 0.025 (0.030). Measured in claims per \$1,000 of insurance in force, the insurer-level relation is stronger: 2.08 (0.65) per excess death per 1,000, 1.26 (0.68) with 1912–1914 weights, and 3.21 (1.29) without the five most affected states. The placebo coefficient of 0.90 (0.67) indicates, however, that part of this relation predates the pandemic, and even taken at face value it implies that geographic exposure accounts for a small share of the variation in insurers’ claims. The pandemic therefore supports the interpretation of Section 6. The largest geographically uneven mortality shock of the period raised claims where it was severe, but the differences in exposure that geographic diversification eliminated were small relative to the other sources of variation in the claims of insurers of a given size.¹⁷

8 Conclusion

This paper documents how American life insurance became a national industry between 1888 and 1940 and uses that expansion to examine whether large intermediaries are stable because they are large or because they are geographically diversified. Insurers became national through many small entries, whose returns were highly skewed and could not be predicted when the entries were made. Entered markets opened at small scale and at lower prices, grew with tenure, and by an insurer’s twentieth year supplied most of its premium. Because insurers grew by entering markets, the largest insurers became the broadest, and a cross-section of insurers cannot reveal which of the two made

¹⁷Online Appendix Section B reports the power of the insurer-level test.

them stable.

Geographic expansion was costly. Before 1920, insurers that had expanded across many states paid more claims per dollar of insurance in force than insurers of the same size writing in the same markets. The difference was a property of the firm rather than of the markets it entered: it appears in the firm-wide level of the exact decomposition, in the insurers' own death claims, and in the headquarters state and the markets an insurer already served. It did not arise as a discrete deterioration at the time an insurer expanded, so it measures the cost associated with the breadth an insurer attained rather than the effect of a particular entry. It declined as the selection of risks was standardized across the industry and had disappeared by the late 1920s. Limited managerial attention to the selection of risks is the interpretation most consistent with this evidence. The pricing cost of expansion lay in the composition of an insurer's business, in the small local shares it held in entered markets.

Geographic breadth did not make insurers of a given size more stable. Theories of intermediation locate the advantage of a large intermediary in the pooling of idiosyncratic risk (Diamond, 1984; Diamond and Dybvig, 1983); for life insurers that advantage came from the number of risks held, not from the number of places in which they were held. Geographic diversification stabilizes an intermediary in proportion to the correlation of its risks within a place. Deaths among a state's insured lives were weakly correlated, and even the 1918–1919 influenza pandemic, the largest geographically uneven mortality shock of the period, produced differences in exposure across insurers that were small relative to other sources of variation in their claims. The local credit risks of banks are more strongly correlated, and geographic expansion reduces bank risk (Goetz et al., 2016). The conclusions concern underwriting risk, and whether geography mattered on the asset side of insurers' balance sheets remains an open question.

References

- Abowd, John M., Francis Kramarz, and David N. Margolis**, “High Wage Workers and High Wage Firms,” *Econometrica*, 1999, 67 (2), 251–333.
- Agarwal, Sumit and Robert Hauswald**, “Distance and Private Information in Lending,” *Review of Financial Studies*, 2010, 23 (7), 2757–2788.
- Albornoz, Facundo, Héctor F. Calvo Pardo, Gregory Corcos, and Emanuel Ornelas**, “Sequential Exporting,” *Journal of International Economics*, 2012, 88 (1), 17–31.
- Andrews, M. J., L. Gill, T. Schank, and R. Upward**, “High Wage Workers and Low Wage Firms: Negative Assortative Matching or Limited Mobility Bias?,” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2008, 171 (3), 673–697.
- Arthi, Vellore, Gary Richardson, and Mark Van Orden**, “Ordinary Lives: Insurance and Savings in America, 1861 to 1941,” 2025. Working paper.
- Association of Life Insurance Medical Directors and Actuarial Society of America**, *Medico-Actuarial Mortality Investigation*, New York: Association of Life Insurance Medical Directors and Actuarial Society of America, 1912–1914. Five volumes.
- Berger, Philip G. and Eli Ofek**, “Diversification’s Effect on Firm Value,” *Journal of Financial Economics*, 1995, 37 (1), 39–65.
- Berry-Stölzle, Thomas R., Andre P. Liebenberg, Joseph S. Ruhland, and David W. Sommer**, “Determinants of Corporate Diversification: Evidence from the Property–Liability Insurance Industry,” *Journal of Risk and Insurance*, 2012, 79 (2), 381–413.
- Bouk, Dan**, *How Our Days Became Numbered: Risk and the Rise of the Statistical Individual*, Chicago: University of Chicago Press, 2015.
- Boyd, John H. and Edward C. Prescott**, “Financial Intermediary-Coalitions,” *Journal of Economic Theory*, 1986, 38 (2), 211–232.
- Brainerd, Elizabeth and Mark V. Siegler**, “The Economic Effects of the 1918 Influenza Epidemic,” *CEPR Discussion Paper No. 3791*, 2003.
- Campa, José Manuel and Simi Kedia**, “Explaining the Diversification Discount,” *Journal of Finance*, 2002, 57 (4), 1731–1762.
- Campbell, John Y., Martin Lettau, Burton G. Malkiel, and Yexiao Xu**, “Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk,” *Journal of Finance*, 2001, 56 (1), 1–43.
- Carlson, Mark and Kris James Mitchener**, “Branch Banking, Bank Competition, and Financial Stability,” *Journal of Money, Credit and Banking*, 2006, 38 (5), 1293–1328.
- Carter, Susan B., Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright, eds**, *Historical Statistics of the United States: Earliest Times to the Present*, millennial ed., New York: Cambridge University Press, 2006.
- Che, Xin and Andre P. Liebenberg**, “Effects of Business Diversification on Asset Risk-Taking: Evidence from the U.S. Property-Liability Insurance Industry,” *Journal of Banking and Finance*, 2017, 77, 122–136.
- Davis, Lance E.**, “The Investment Market, 1870–1914: The Evolution of a National Market,” *Journal of Economic History*, 1965, 25 (3), 355–399.
- Demsetz, Rebecca S. and Philip E. Strahan**, “Diversification, Size, and Risk at Bank Holding Companies,” *Journal of Money, Credit and Banking*, 1997, 29 (3), 300–313.
- Diamond, Douglas W.**, “Financial Intermediation and Delegated Monitoring,” *Review of Economic Studies*, 1984, 51 (3), 393–414.
- **and Philip H. Dybvig**, “Bank Runs, Deposit Insurance, and Liquidity,” *Journal of Political Economy*, 1983, 91 (3), 401–419.

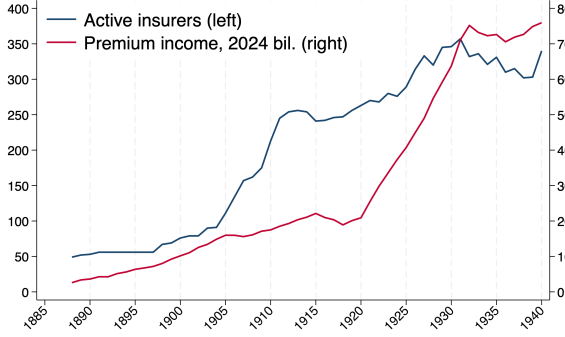
- Foley-Fisher, Nathan, Borghan Narajabad, and Stephane Verani**, “Self-Fulfilling Runs: Evidence from the U.S. Life Insurance Industry,” *Journal of Political Economy*, 2020, 128 (9), 3520–3569.
- Goetz, Martin R., Luc Laeven, and Ross Levine**, “Does the Geographic Expansion of Banks Reduce Risk?,” *Journal of Financial Economics*, 2016, 120 (2), 346–362.
- Janas, Pawel**, “Integration Without Decentralization: The Financial Geography of U.S. Life Insurance, 1880 - 1940,” 2026. Working paper.
- Jensen, Michael C.**, “Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers,” *American Economic Review*, 1986, 76 (2), 323–329.
- Johnston, Louis and Samuel H. Williamson**, “What Was the U.S. GDP Then?,” Measuring-Worth 2026. Accessed: 2026-07-20.
- Keller, Morton**, *The Life Insurance Enterprise, 1885–1910: A Study in the Limits of Corporate Power*, Cambridge, MA: Belknap Press of Harvard University Press, 1963.
- Koijen, Ralph S. J. and Motohiro Yogo**, “The Fragility of Market Risk Insurance,” *Journal of Finance*, 2022, 77 (2), 815–862.
- Lang, Larry H. P. and René M. Stulz**, “Tobin’s q , Corporate Diversification, and Firm Performance,” *Journal of Political Economy*, 1994, 102 (6), 1248–1280.
- Liberti, José María and Atif R. Mian**, “Estimating the Effect of Hierarchies on Information Use,” *Review of Financial Studies*, 2009, 22 (10), 4057–4090.
- Liebenberg, Andre P. and David W. Sommer**, “Effects of Corporate Diversification: Evidence from the Property–Liability Insurance Industry,” *Journal of Risk and Insurance*, 2008, 75 (4), 893–919.
- Murphy, Sharon Ann**, *Investing in Life: Insurance in Antebellum America*, Baltimore: Johns Hopkins University Press, 2010.
- North, Douglass C.**, “Life Insurance and Investment Banking at the Time of the Armstrong Investigation of 1905–1906,” *Journal of Economic History*, 1954, 14 (3), 209–228.
- Ransom, Roger L and Richard Sutch**, “Tontine insurance and the Armstrong investigation: a case of stifled innovation, 1868–1905,” *The Journal of Economic History*, 1987, 47 (2), 379–390.
- Rogers, Oscar H. and Arthur Hunter**, “The Numerical Method of Determining the Value of Risks for Insurance,” *Transactions of the Actuarial Society of America*, 1919, 20, 273–300.
- Rose, Jonathan D**, “Financial crises at insurance companies: learning from the demise of the National Surety Company during the Great Depression,” *Financial History Review*, 2017, 24 (3), 239–264.
- Ruhl, Kim J. and Jonathan L. Willis**, “New Exporter Dynamics,” *International Economic Review*, 2017, 58 (3), 703–726.
- Snowden, Kenneth A.**, “Mortgage Securitization in the United States: Twentieth Century Developments in Historical Perspective,” in Michael D. Bordo and Richard Sylla, eds., *Anglo-American Financial Systems: Institutions and Markets in the Twentieth Century*, Burr Ridge, IL: Irwin, 1995.
- Stalson, J. Owen**, *Marketing Life Insurance: Its History in America*, Cambridge, MA: Harvard University Press, 1942.
- Stein, Jeremy C.**, “Information Production and Capital Allocation: Decentralized versus Hierarchical Firms,” *Journal of Finance*, 2002, 57 (5), 1891–1921.
- U.S. Bureau of the Census**, “Mortality Statistics,” Technical Report, U.S. Government Printing Office, Washington, DC 1920. Annual volumes for 1918 and 1919.
- Villalonga, Belén**, “Diversification Discount or Premium? New Evidence from the Business Information Tracking Series,” *Journal of Finance*, 2004, 59 (2), 479–506.

- Williamson, Samuel H.**, “The Annual Consumer Price Index for the United States, 1774–Present,” MeasuringWorth 2026. Accessed: 2026-07-20.
- Winton, Andrew**, “Don’t Put All Your Eggs in One Basket? Diversification and Specialization in Lending,” Working Paper 00-16, Wharton Financial Institutions Center 1999.

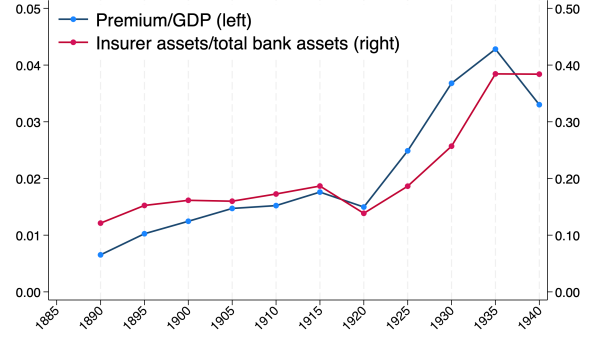
9 Figures

Figure I: Life Insurance Industry, 1888–1940

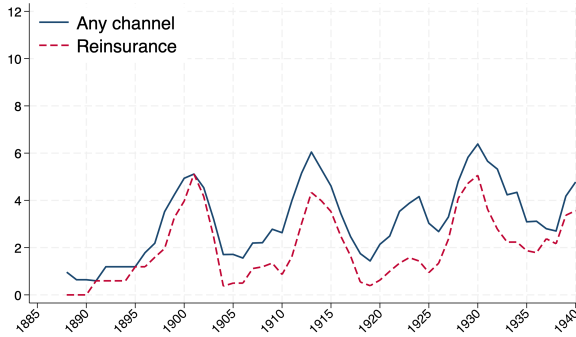
Panel A. Number of insurers and premium income



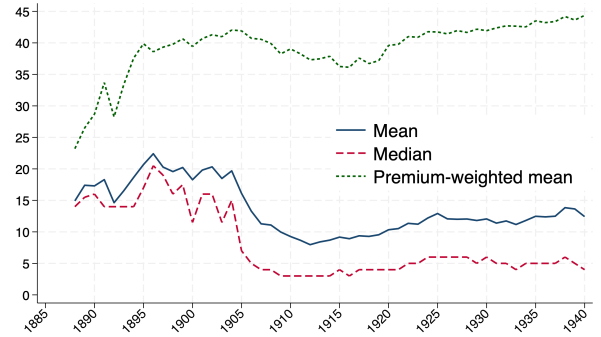
Panel B. Assets and premium shares



Panel C. Exit rates, % (three-year average)

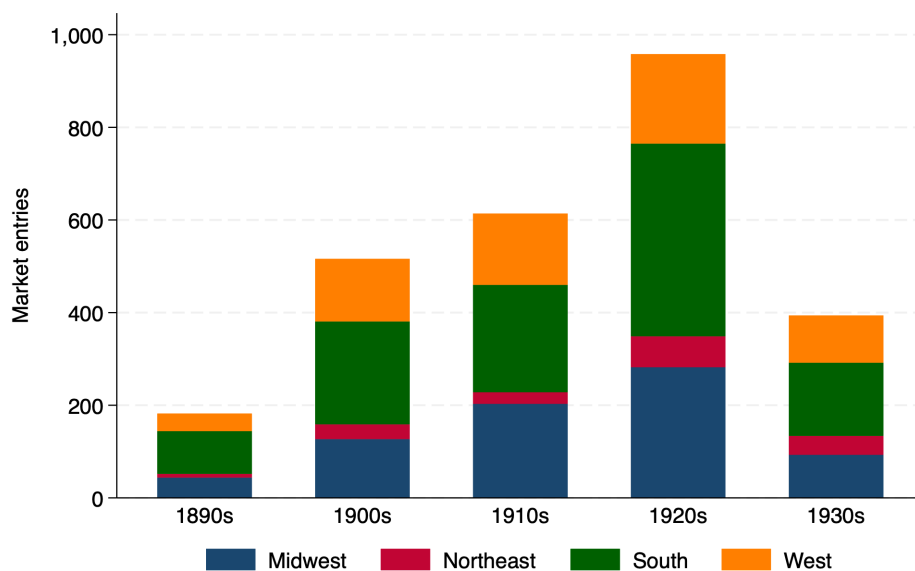


Panel D. States per company



Notes. Nominal GDP and the price deflator are from MeasuringWorth (Williamson (2026); Johnston and Williamson (2026)) and commercial-bank assets from *Historical Statistics of the United States* (Carter et al., eds, 2006). Panel A plots the number of active legal-reserve life insurers (left axis) and aggregate premium income in 2024 dollars (right axis). Panel B plots premium income as a share of GDP (left axis) and life-insurer assets as a share of total commercial-bank assets (right axis). Panel C plots the annual rate at which active insurers exit, as a three-year moving average ($t - 1, t, t + 1$), through any channel and through reinsurance alone. Panel D plots the number of states in which a firm writes premium, as the cross-firm mean, the median, and the premium-weighted mean.

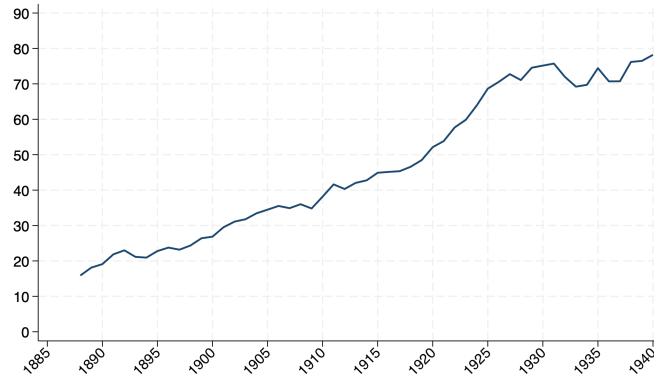
Figure II: Market Entries by Decade and Region



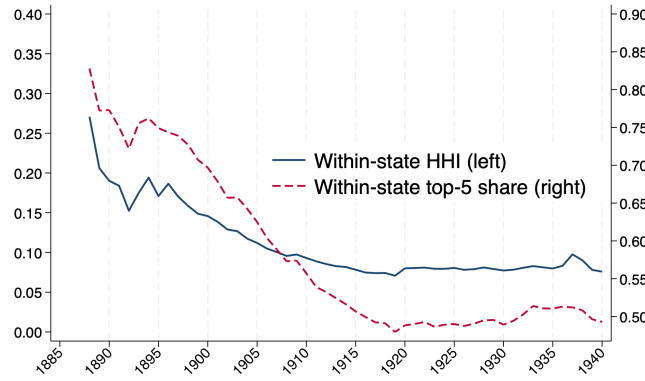
Notes. Number of market entries by decade and by the census region of the entered state, ordinary insurance, 1890–1940. An entry is a firm-state’s first observed year with positive premium, after the insurer’s first year and after 1889, in which new insurance written exceeds half of insurance in force and premium does not exceed the insurer’s median state premium that year (Section 3.3).

Figure III: Market Structure, 1888–1940

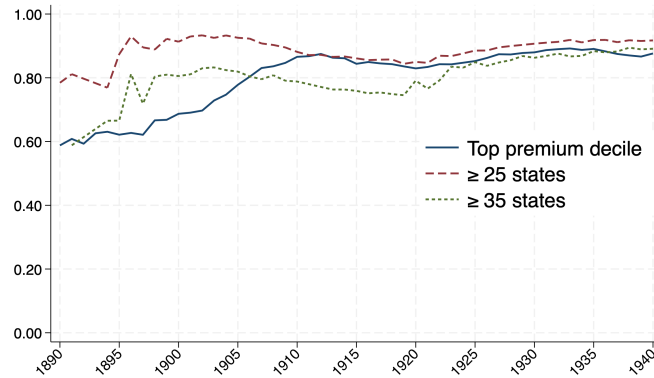
Panel A. Mean insurers per state



Panel B. Local concentration



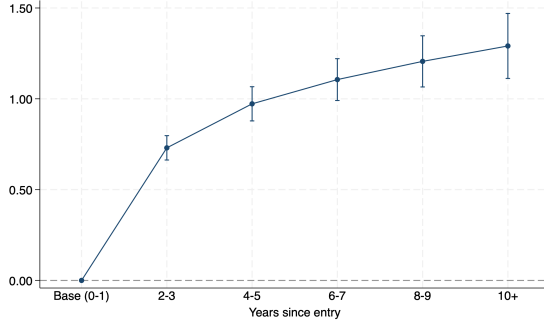
Panel C. National premium share of broad insurers



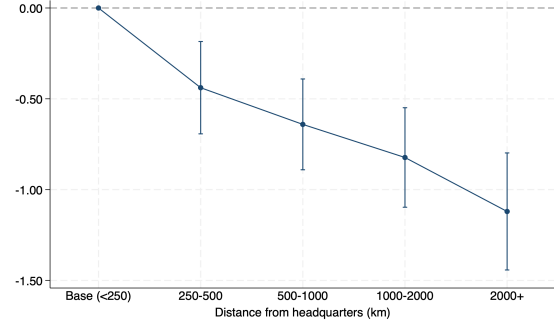
Notes. Panel A plots the average number of insurers operating in a state. Panel B plots the within-state Herfindahl index of premiums (left axis) and the premium share of the five largest firms in a state (right axis). Panel C plots the share of national premiums collected by firms operating in at least twenty-five states, in at least thirty-five states, and in the top decile of the premium distribution.

Figure IV: The Scale and Claims Experience of Entered Markets

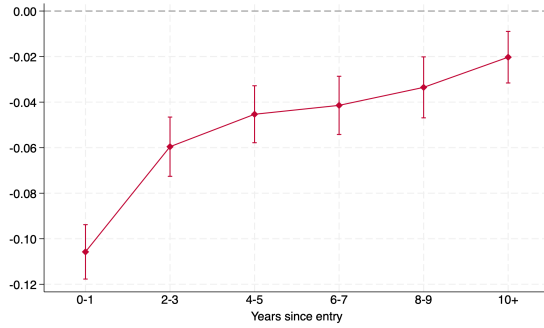
Panel A. Log premium by years since entry, relative to the market's first two years



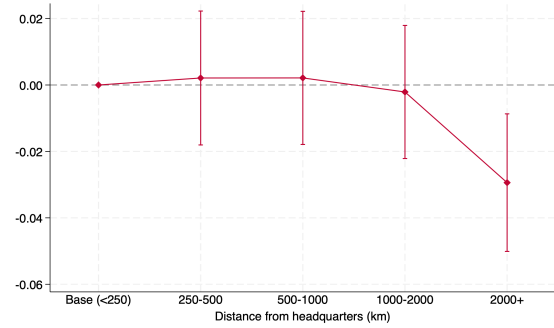
Panel B. Log premium of entered markets by distance, relative to entered markets within 250 km



Panel C. Loss ratio by years since entry, relative to the insurer's founding states



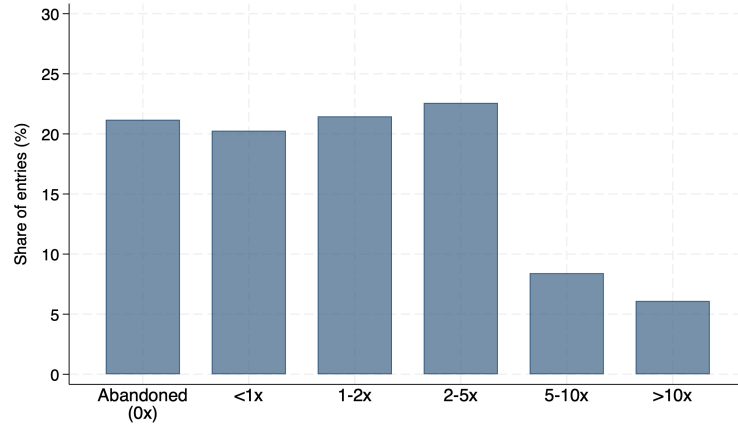
Panel D. Loss ratio of entered markets by distance, relative to entered markets within 250 km



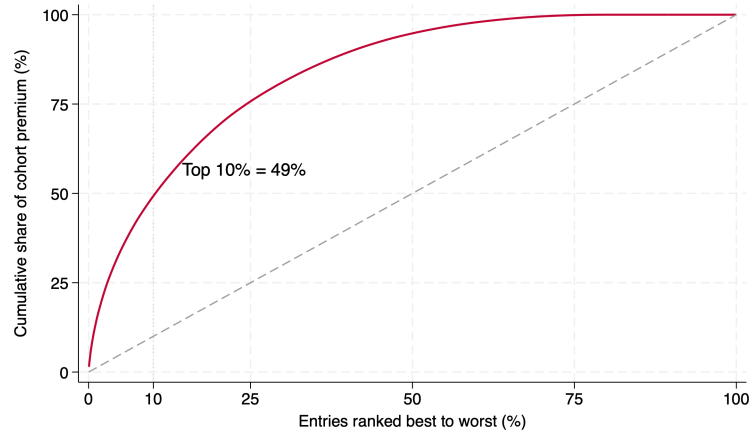
Notes. Ordinary insurance, 1888–1940. Each panel plots the coefficients from a regression of the outcome on indicator variables for bins of the argument shown, with indicators for the other argument included as controls; the omitted bin is the reference group named in the panel title. Panels A and B: log premium. Panel A uses two-year bands of tenure with firm-state and firm $\times$ year fixed effects, so that each market's average size is removed and the profile follows a single market as it ages rather than a cohort effect. Panel B restricts to entered markets and absorbs firm $\times$ year fixed effects alone, since distance does not vary within a firm-state. Panels C and D: the loss ratio, claims over premiums, with firm $\times$ year fixed effects; Panel C uses two-year bands of tenure, and Panel D restricts to entered markets. No-claim state-years are retained and the top 3 percent of loss ratios is trimmed. Distance bands are under 250, 250–500, 500–1,000, 1,000–2,000, and over 2,000 kilometers from the headquarters city to the state's centroid. Standard errors are clustered by firm and bars are 95 percent confidence intervals.

Figure V: Market Entry Payoffs

Panel A. Outcome distribution

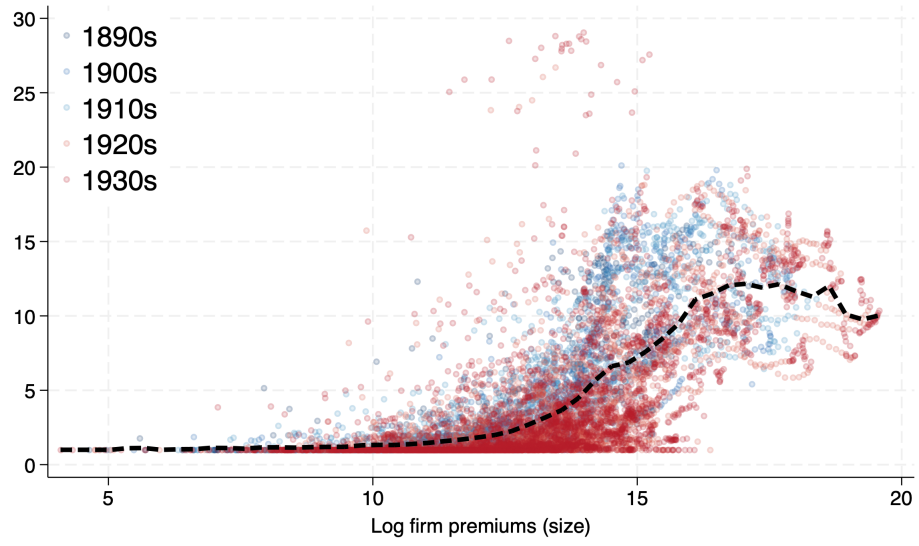


Panel B. Lorenz curve



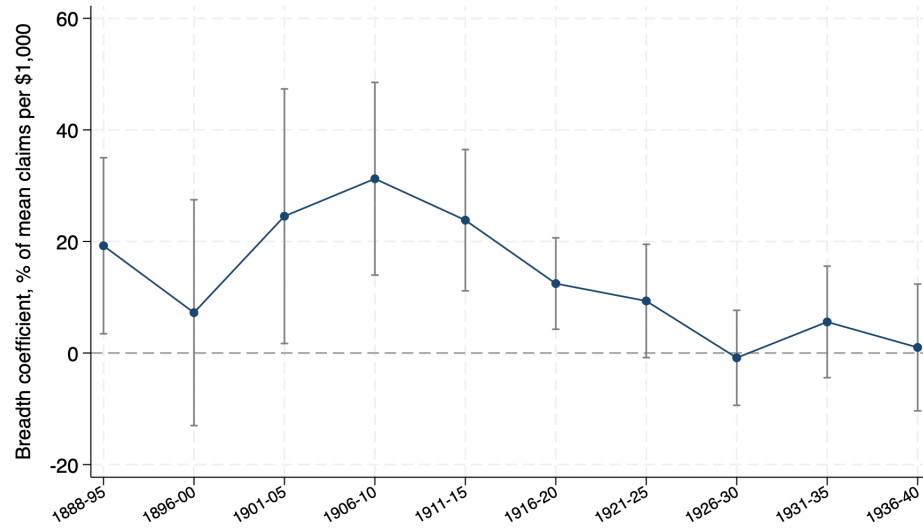
Notes. The figure describes the ten-year outcomes of true market entries, ordinary insurance, 1890–1940. The sample is restricted to entries whose firm survives at least ten years past entry. An entry is *abandoned* if the firm stops writing in that state within ten years while still operating; its payoff is zero. For surviving entries the payoff multiple is the market's premium about ten years after entry (averaged over years nine through eleven) relative to its premium in the second year after entry, divided by the firm's own total-premium multiple over the same window. Panel A groups entries by this multiple; Panel B plots the cumulative share of the cohort's year-ten premium against entries ranked from best to worst.

Figure VI: Geographic Diversification and Firm Size



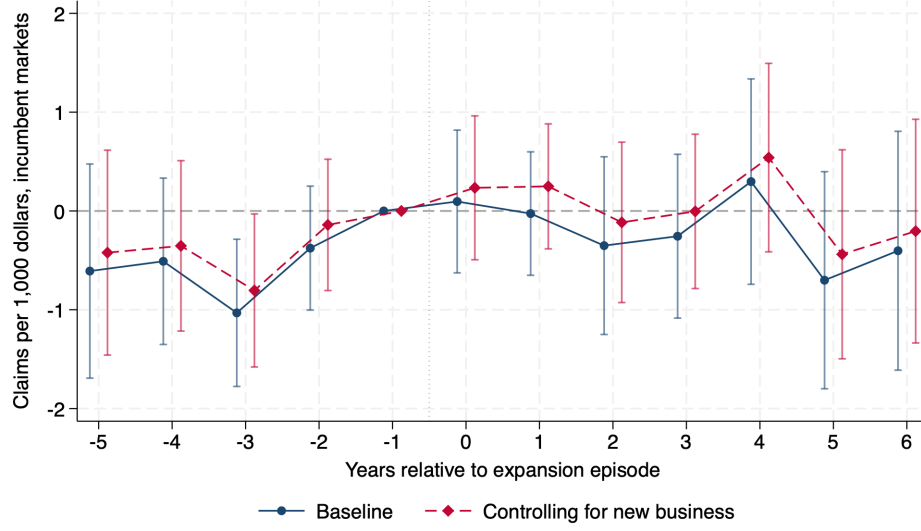
Notes. This figure plots the effective number of states $N_{\text{eff}} = 1/\sum_s w_s^2$ (the inverse Herfindahl) against log firm premium. Each firm-year is colored by decade.

Figure VII: The Claims Differential of Broad Insurers over Time



Notes. Coefficient on breadth in regressions of the claims level term of the exact decomposition (Table VIII) on breadth and a cubic in size, with headquarters-region-by-year effects, estimated separately for each period (1888–1895, then five-year periods from 1896–1900 to 1936–1940) on the 1888–1940 sample, in percent of the period’s mean claims per \$1,000 of insurance in force. Breadth and size are standardized over 1888–1940. Bars are 95 percent confidence intervals based on standard errors clustered by firm.

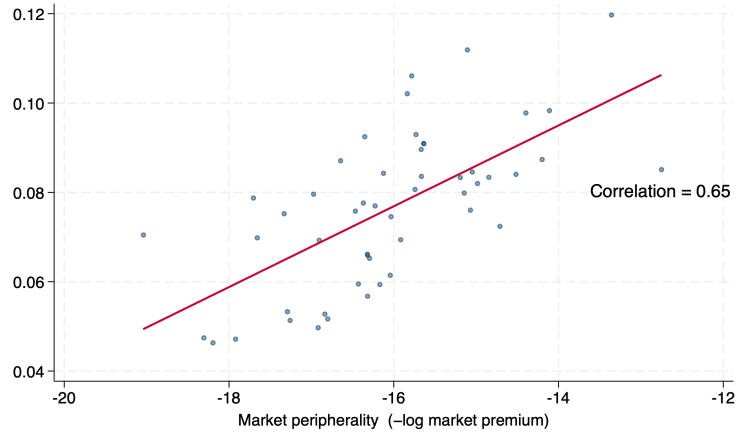
Figure VIII: Claims in Incumbent Markets around Expansion Episodes



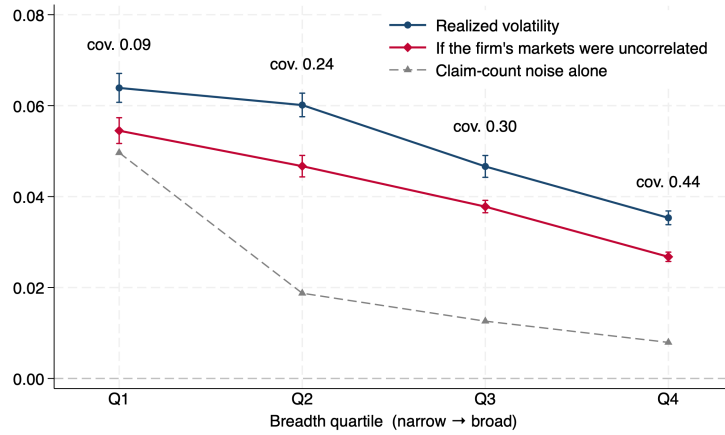
Notes. Coefficients β_k of equation (5.4), claims per \$1,000 of insurance in force in the incumbent markets of insurers experiencing an expansion episode, relative to the same states' markets of insurers that had not yet expanded, by year relative to the episode; year -1 is omitted. An expansion episode is the first year after an insurer's first appearance in which it enters three or more states; the sample consists of the incumbent markets of the 97 treated insurer-episodes and of control insurers that had not yet expanded, over years -5 to 6 , with market-by-stack and state-by-year-by-stack fixed effects, a cubic in size, local share, and tenure indicators. Pooled over years 0 to 6 , the baseline coefficient is -0.17 (standard error 0.37), against a mean of 6.60 before the episode; the leads, years -5 to -2 , are not jointly significant ($p = 0.12$). Bars are 95 percent confidence intervals based on standard errors clustered by firm.

Figure IX: The Limits of Geographic Pooling

Panel A. Loss volatility vs. market size



Panel B. Realized volatility and the own-market benchmark



Notes. The sample is ordinary-insurance business, 1888–1940. Panel A plots each state’s loss-ratio volatility (the standard deviation of its premium-weighted loss ratio across years, for states observed at least five years) against market peripherality, measured as the negative log of average state premium, with an ordinary-least-squares fit. Panel B uses five-year windows of consecutive firm-years for firms writing in at least two states and reports, by breadth quartile (narrow to broad), medians across windows of the realized volatility of the firm’s aggregate loss ratio, $\sqrt{R}$; of the volatility it would have if its markets were uncorrelated, $\sqrt{B}$ with $B = \sum_s \bar{w}_s^2 \text{Var}(lr_s)$ computed from the firm’s own state loss ratios in the same window; and of the volatility implied by claim-count sampling alone, where policy counts are reported. Bars are 95 percent intervals for the medians. Labels give the median covariance share $1 - B/R$ (Online Appendix Table [IA.25a](#)). Market cells that fail the validity screen are excluded.

10 Tables

Table I: Summary Statistics

	Mean	SD	p25	p50	p75	N
<i>Panel A: Firm-state-year</i>						
Premiums (k)	471.02	2,288.12	12.89	62.13	257.25	146,650
Losses (k)	156.54	735.66	2.00	16.54	80.59	146,650
Insurance written (k)	2,679.93	11,250.97	107.00	489.38	1,745.50	146,649
Firm-state market share	0.02	0.04	0.00	0.00	0.02	146,650
Firm-state exit	0.03	0.18	0.00	0.00	0.00	146,650
<i>Panel B: Firm-year</i>						
Total premiums (k)	6,711.39	35,842.31	168.52	574.29	2,037.75	10,287
Total losses (k)	2,231.58	11,797.60	28.15	127.93	603.25	10,287
Insurance written (m)	38.15	193.77	1.85	5.44	16.83	10,287
Number of states	14.16	18.91	2.00	6.00	20.00	10,287
Breadth	4.43	4.75	1.10	2.15	6.01	10,287
Loss ratio	0.26	0.16	0.14	0.22	0.33	10,081
Loss-ratio volatility (5-yr)	0.07	0.06	0.03	0.05	0.09	9,433
Exit: failure (10y)	0.02	0.14	0.00	0.00	0.00	10,287
Exit: merger (10y)	0.04	0.19	0.00	0.00	0.00	10,287
Exit: reinsurance (10y)	0.13	0.34	0.00	0.00	0.00	10,287
Exit: acquisition (10y)	0.02	0.13	0.00	0.00	0.00	10,287
<i>Panel C: NY accounting sample, firm-year</i>						
Surplus / total liabilities	0.12	0.12	0.05	0.08	0.15	2,261
Mortgages / assets	0.32	0.19	0.18	0.31	0.45	2,390
Real estate / assets	0.07	0.08	0.01	0.04	0.10	2,390
(Stocks + bonds) / assets	0.41	0.21	0.27	0.39	0.54	2,390
Policy loans / assets	0.08	0.07	0.01	0.08	0.14	2,342
(Surrenders + lapses) / in-force	0.08	0.06	0.04	0.06	0.09	2,253
Bond interest income / assets	0.04	0.01	0.04	0.04	0.05	2,294
Accounting loss ratio	0.41	0.19	0.30	0.38	0.50	2,385
<i>Panel D: Spectator company accounts, firm-year</i>						
Premium income (k)	4,518.63	18,702.19	132.27	456.94	1,805.42	5,866
Death claims paid (k)	1,297.57	4,711.73	23.61	92.04	518.49	5,561
Insurance in force (m)	104.60	364.95	3.36	12.02	49.04	5,868
Policies in force (k)	51.00	200.48	2.12	7.22	26.38	5,840
New insurance issued (m)	18.28	60.38	1.26	3.43	10.66	5,840
Deaths (policies)	490.31	1,710.45	13.00	50.00	246.00	4,829
Average policy in force (\$)	1,900.53	654.54	1,511.59	1,851.38	2,277.71	5,724
Death claims per \$1,000 in force	11.91	20.64	4.14	7.07	11.74	5,296
Deaths per 1,000 policies	7.22	4.03	4.18	6.36	9.47	4,725
Actual / expected mortality	0.60	0.21	0.45	0.57	0.72	3,137
Medical fees per \$1,000 new insurance	3.42	2.97	2.16	2.81	3.64	4,195
Lapses per \$1,000 in force	130.04	99.73	56.57	107.54	174.09	4,845
Endowment share of in force	0.15	0.14	0.05	0.10	0.20	4,704

Notes. Panel A reports firm-state-year observations, all lines of business, for U.S. legal-reserve life insurers in the *Spectator Insurance Year Book*, 1888–1940, after the validity screen; Panel B reports firm-year observations from the same source, excluding firm-years with no recorded losses; Panel C reports firm-year observations from the financial statements of the New York sample; Panel D reports firm-year observations, 1886–1926, from the company accounts and gain-and-loss exhibits of the *Insurance Year Book* for the insurers linked to the panel. Dollar amounts and policy counts are in thousands (k), except the rows marked (m), whose medians exceed a million dollars, which are in millions. The firm-year exit rows report the share of firm-years followed within ten years by an exit through the indicated channel. In Panel D the figures are company-wide and include the industrial departments of combination companies; rates per \$1,000 and per 1,000 policies use the mean of the current and previous year-end insurance in force, and the rates and ratios are trimmed at the 1st and 99th percentiles.

Table II: Sources of the Growth of Industry Premium

<i>Panel A. Real change in industry premium and its sources (percent of premium at the start of the decade)</i>						
Decade	Premium at start (2024 \$bn)	Total change	Incumbent markets	Entered markets	Insurers first appearing	Abandoned markets
1890–1900	3.42	137.2	98.4	32.1	9.0	-2.4
1900–1910	8.11	70.3	56.5	6.7	13.3	-6.3
1910–1920	13.81	16.9	10.9	4.6	6.1	-4.7
1920–1930	16.14	180.9	171.0	6.7	6.8	-3.6
1930–1940	45.33	16.5	18.0	1.6	2.9	-6.0
<i>Panel B. Premium of entered markets by census region of the entered state (percent)</i>						
Decade	Northeast	Midwest	South	West		
1890–1900	17.4	26.5	41.5	14.5		
1900–1910	3.8	21.2	51.3	23.7		
1910–1920	2.0	38.9	43.6	15.5		
1920–1930	9.1	36.5	36.5	17.9		
1930–1940	18.4	36.2	33.3	12.1		

Notes. Ordinary premium of all insurers in the analysis sample, after the validity screen, in the first and last year of each decade. Panel A reports the real change in industry premium over the decade (total change) and divides it into four sources, each in percent of industry premium in the first year of the decade: the change in premium of markets an insurer served in both years (incumbent markets); the premium in the last year of markets entered during the decade by insurers active in its first year (entered markets); the premium of insurers not active in the first year (insurers first appearing); and minus the first-year premium of markets an insurer no longer served in the last year (abandoned markets). The four sources sum to the total change. Panel B reports the distribution of the premium of entered markets across the census regions of the entered states. Premium is converted to constant 2024 dollars with the consumer price index used in Figure I; premium at the start of the decade is in billions.

Table III: The Geographic Structure of the Industry

Year	Headquarters-state share		Out-of-state share (mean over states)	Effective number of states	
	Median	Premium-weighted		Median	Premium-weighted
1890	0.28	0.22	0.94	4.07	8.61
1900	0.33	0.22	0.93	4.58	10.49
1910	0.80	0.22	0.89	1.43	11.31
1920	0.61	0.22	0.87	1.93	10.92
1930	0.65	0.22	0.89	1.95	9.89
1940	0.58	0.20	0.90	2.07	10.85

Notes. Ordinary insurance, after the validity screen. The headquarters-state share is the share of an insurer's premium collected in the state in which it was headquartered, as the median across insurers and weighted by premium. The out-of-state share is the share of a state's premium written by insurers headquartered in other states, averaged across states. The effective number of states is $N_{\text{eff}} = 1/\sum_s w_s^2$, as the median across insurers and weighted by premium.

Table IV: Premium by Market Vintage

Market vintage	Firm age (years)			
	5	10	15	20
Founding states	68	57	50	43
Entered 1–5 yrs after	32	34	35	38
Entered 6–10 yrs after	—	9	8	10
Entered 11–15 yrs after	—	—	7	6
Entered 16–20 yrs after	—	—	—	2
Post-founding entries (all)	32	43	50	57

Notes. The table reports the share of a firm’s ordinary premium, at a given firm age, coming from the states it was founded with versus the states it entered afterward, grouped by how many years after founding it entered them. Shares are premium-weighted. A market’s vintage is its entry age, the year the firm first wrote business in the state minus the firm’s founding year. To measure founding cleanly for firms already operating when the panel opens, each firm’s pre-existing markets are removed: any firm-state present in the first sample year is treated as null throughout, and founding is reset to the firm’s first market observed thereafter; firms formed during the panel are unaffected.

Table V: Time to National Scope, by Cohort of First Appearance

Cohort of first appearance	Insurers	Percent reaching within 10 years			Median years, insurers that reach it		
		3 states	5 states	10 states	3 states	5 states	10 states
1890–1899	31	64.0	52.2	20.0	2	3	6
1900–1909	147	35.9	15.0	4.1	8	11	17
1910–1919	182	29.2	18.6	5.9	6	6	6
1920–1929	190	41.9	22.4	12.0	2	2	5

Notes. Insurers in the analysis sample first observed between 1890 and 1929, grouped by the decade of their first appearance. For each threshold of 3, 5, and 10 effective states (N_{eff}), the first three columns report the percentage of insurers that reached it within ten years of their first appearance, among insurers observed for at least ten years or reaching it earlier; the last three report the median number of years from first appearance to the threshold among insurers that reached it.

Table VI: Distance and Local Pricing Power
Panel A. The distance penalty and the margins of the loss ratio

	Loss ratio		Loss-ratio margins		Premium yield	
	(1)	(2)	Pr(no claims)	Loss ratio > 0	(5)	(6)
Log distance	-0.039*** (0.004)	-0.029*** (0.003)	0.083*** (0.007)	-0.016*** (0.004)	-0.739*** (0.087)	-0.583*** (0.089)
State tenure		0.002*** (0.000)				0.017 (0.018)
$E[Y]$	0.276	0.276	0.176	0.337	30.283	30.283
Observations	116513	116513	116513	95236	113197	113197
R2	0.30	0.31	0.33	0.28	0.43	0.43

Panel B. Local pricing power

	Premium yield	Loss ratio	Premium yield
Within-state share	17.13** (7.67)	-0.18** (0.08)	16.79*** (3.53)
Log distance			-0.64*** (0.09)
Log distance $\times$ share			-2.81 (2.46)
$E[Y]$	30.291	0.275	30.283
Observations	114170	117598	113197
R2	0.42	0.29	0.43

Panel C. Markets weighted by their contribution to the insurer's totals

	Loss ratio		Claims per 1,000		Premium yield	
	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted
Log distance	-0.039*** (0.004)	-0.015*** (0.004)	-1.336*** (0.122)	-0.661*** (0.156)	-0.739*** (0.087)	-0.588*** (0.128)
$E[Y]$	0.276	0.331	8.334	10.583	30.283	31.964
Observations	116513	116513	114250	114250	113197	113197
R2	0.30	0.44	0.32	0.48	0.43	0.60

Notes. The unit of observation is the firm-state-year, ordinary insurance, 1888–1940. All columns include firm $\times$ year fixed effects and standard errors are clustered by firm and by state-year. Distance is the great-circle distance in kilometers from the firm's headquarters city to the state's centroid in logs; within-state share is the firm's premium in a state divided by all firms' premium in that state that year; and state tenure is years since the firm entered the state. The loss ratio is claims over premiums, premium yield is premiums per \$1,000 of insurance in force, and claims per \$1,000 is claims over insurance in force. $E[Y]$ reports the mean of each column's outcome on that column's own estimation sample. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table VII: The Pricing Cost of Entered Markets

	1888–1920	1921–1940	1888–1940
<i>Panel A. The pricing cost of entered markets</i>			
Premium per \$1,000, entered minus original markets	-0.40*	-0.51***	-0.47***
	(0.22)	(0.17)	(0.14)
conditional on tenure	-0.37*	-0.28	-0.31**
	(0.22)	(0.18)	(0.15)
conditional on local share	0.76***	0.27	0.48***
	(0.24)	(0.18)	(0.15)
Premium per \$1,000 in entered markets	31.8	30.3	30.6
Forgone premium / premium earned in entered markets	0.013	0.017	0.015
95% band	[-0.001, 0.026]	[0.006, 0.029]	[0.006, 0.025]
Entered markets' share of discounted premium	0.215	0.276	0.254
Market-years (firm-years holding both kinds)	45,596	64,570	110,150
<i>Panel B. Upper bound: the level term treated as a cost of expansion</i>			
Loss-ratio penalty per log point of N^{eff}	0.058	0.019	0.032
Break-even loading on premium, m^*	0.149	0.048	0.082
95% band from the penalty's s.e.	[0.091, 0.207]	[-0.003, 0.100]	[0.042, 0.122]
adding the commission penalty	0.202	0.103	0.136
Share of firms for which expansion paid, loading 10%	0.33	0.94	0.87
loading 15%	0.82	0.95	0.92
loading 20%	0.88	0.97	0.94
Firms	299	405	508

Notes. Panel A: market cells with positive insurance in force; the gap is the coefficient on an indicator for markets entered after the firm's first year in a regression of premium per \$1,000 (trimmed at the 1st and 99th percentiles) with firm-by-year and state-by-year effects, weighting markets by their share of the firm's insurance in force, on firm-years holding both kinds of market; the conditional rows add tenure bins or local-share quintiles. Forgone premium is minus the gap times the insurance in force of entered markets, discounted at 4 percent to the firm's first year, relative to the discounted premium of entered markets; the band uses the gap's 95 percent confidence interval. Panel B is an upper bound that treats the whole level term as caused by expansion: for each firm, the cost is the loss-ratio penalty per log point of the effective number of states times the change in $\log N^{\text{eff}}$ since the firm's first year in the panel, times its total premium; the benefit is a loading times the premium it wrote in markets entered after that year, all of which is treated as incremental; both are discounted at 4 percent and summed over firms, and the break-even loading equates them. By period, the penalty is estimated within the period. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table VIII: Breadth and Firm Outcomes: Preferred Specification and Exact Decomposition

	Firm aggregate β^{γ}	Level β^{ϕ}	Markets β^{α}	Position β^{δ}	N
<i>Panel A. 1888–1920 (% of the outcome mean)</i>					
Loss ratio	22.6*** (4.5)	25.8*** (4.5)	-0.9 (0.9)	-2.3*** (0.9)	4,162
Claims per \$1,000 in force	18.7*** (4.4)	21.8*** (4.3)	-1.9 (1.2)	-1.2 (1.1)	4,162
Premium per \$1,000 in force	-5.6*** (1.3)	-4.7*** (1.3)	-0.8** (0.4)	-0.1 (0.2)	4,162
<i>Panel B. 1921–1940</i>					
Loss ratio	6.8* (3.7)	5.6 (4.3)	1.7*** (0.6)	-0.4 (2.4)	5,733
Claims per \$1,000 in force	3.8 (3.6)	2.5 (3.9)	1.8*** (0.5)	-0.5 (1.9)	5,733
Premium per \$1,000 in force	-3.1*** (1.0)	-2.5*** (1.0)	-0.0 (0.1)	-0.5*** (0.1)	5,733
<i>Panel C. 1888–1940</i>					
Loss ratio	12.2*** (3.0)	12.9*** (3.5)	1.2** (0.6)	-1.9 (1.9)	9,897
Claims per \$1,000 in force	9.6*** (3.0)	10.3*** (3.2)	0.8 (0.6)	-1.6 (1.5)	9,897
Premium per \$1,000 in force	-3.5*** (0.8)	-2.9*** (0.8)	-0.2 (0.2)	-0.4*** (0.1)	9,897
<i>Position including local-share quintiles</i>					
Loss ratio	12.2*** (3.0)	10.6*** (3.4)	1.5*** (0.6)	0.1 (1.9)	9,897
Claims per \$1,000 in force	9.6*** (3.0)	9.9*** (3.2)	0.9 (0.6)	-1.2 (1.5)	9,897
Premium per \$1,000 in force	-3.5*** (0.8)	-1.1 (0.8)	-0.4* (0.3)	-1.9*** (0.2)	9,897
<i>Panel D. 1888–1940, firm aggregate</i>					
	Unconditional	Size cubic	Size and maturity	Logarithms	N
Loss ratio (mean 0.244)	13.6*** (2.3)	12.2*** (3.0)	11.6*** (2.9)	12.2*** (2.7)	9,897
Claims per \$1,000 in force (mean 7.42)	14.4*** (2.3)	9.6*** (3.0)	8.6*** (2.7)	9.0*** (2.7)	9,897
Premium per \$1,000 in force (mean 30.70)	0.2 (0.6)	-3.5*** (0.8)	-3.9*** (0.9)	-3.2*** (0.9)	9,897

Notes. The sample consists of firm-years of ordinary insurance constructed from market cells with positive insurance in force that pass the validity screen of Section 3.3; outcomes are in levels. A single joint trim removes the top percentile of the loss ratio and claims and both tails of premium; zero-loss firm-years are retained. Each cell reports the coefficient on breadth (standardized log effective number of states) in percent of the outcome mean, with the standard error below, conditional on a cubic in standardized log premium; all specifications include region-by-year effects, and standard errors are clustered by firm. Panels A–C estimate equation (5.1) with firm-by-year and state-by-year effects and a position term of tenure bins (separate for markets held in 1888) by distance quintiles, weight markets by their share of premium (loss ratio) or insurance in force (claims, premium), and regress each term of equation (5.2) on breadth; the level, markets, and position columns sum exactly to the firm aggregate. Breadth and size are standardized within each period. The lower block of Panel C adds quintiles of the firm's share of the state's premium to the position term. Panel D: *size and maturity* adds firm age, premium-weighted market tenure, and an indicator for firms present in 1888; *logarithms* uses log outcomes on the firm-years with positive losses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table IX: Breadth and Outcomes within Markets, by Period

	Loss ratio	Claims per \$1,000	Premium per \$1,000
<i>Panel A. Size controls</i>			
1888–1905	28.7*** (6.3)	26.7*** (5.9)	-3.2 (3.1)
1906–1920	18.7*** (5.7)	17.4*** (5.6)	-1.1 (1.2)
1921–1940	4.5 (3.2)	2.7 (3.0)	-2.3** (0.9)
1888–1940	10.7*** (3.1)	8.9*** (3.0)	-2.0*** (0.8)
<i>Panel B. Adding local share, tenure, and the founding-market indicator</i>			
1888–1905	26.7*** (6.5)	24.1*** (6.1)	-3.4 (3.1)
1906–1920	12.8*** (4.9)	11.5** (4.5)	-1.1 (1.3)
1921–1940	5.1* (2.9)	3.6 (2.6)	-2.1** (0.9)
1888–1940	9.9*** (2.7)	8.5*** (2.5)	-1.8** (0.8)
Mean, 1888–1905	0.321	10.12	32.07
Mean, 1906–1920	0.289	9.10	32.09
Mean, 1921–1940	0.252	7.36	28.99
Observations, 1888–1940	120,152	117,853	116,639

Notes. Market cells of ordinary insurance that pass the validity screen. Each cell is the coefficient θ on breadth in equation (5.3), estimated separately for each period and expressed in percent of the outcome's mean in that period, with each market weighted equally. All specifications include state-by-year fixed effects and a cubic in size; Panel B adds the insurer's share of the state's premium, its tenure in the state, and an indicator for markets held in 1888. Breadth, the standardized log effective number of states, and size, the standardized log of total premium, are measured at the firm-year level and standardized over 1888–1940. The loss ratio and claims per \$1,000 of insurance in force are trimmed at the 97th percentile and premium per \$1,000 at the 1st and 97th. Standard errors, clustered by firm and by state-year, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table X: The Incidence of the Claims Differential within an Insurer's Markets

	1888–1905	1906–1920	1921–1940
<i>Panel A. Headquarters state, largest market, and distance</i>			
Breadth (nearest third of other markets)	3.65*** (0.81)	1.30*** (0.44)	0.44** (0.22)
Breadth $\times$ headquarters state	-1.33** (0.67)	-0.41 (0.27)	-0.55*** (0.16)
Breadth $\times$ largest other market (previous year)	-1.64 (1.05)	-1.62*** (0.51)	-1.07*** (0.27)
Breadth $\times$ middle third by distance	-1.15** (0.56)	-0.09 (0.24)	0.10 (0.16)
Breadth $\times$ farthest third by distance	-1.55** (0.77)	0.07 (0.34)	0.09 (0.21)
<i>Panel B. Local share and tenure</i>			
Breadth	3.02*** (0.91)	1.28*** (0.49)	0.31* (0.18)
Breadth $\times$ local share (centered)	-9.35** (3.94)	-2.35 (5.85)	1.94 (4.98)
Breadth $\times$ tenure (decades, centered)	0.52 (0.57)	0.35 (0.28)	0.05 (0.13)
<i>Panel C. Markets outside the Northeast</i>			
Breadth (Northeastern markets, nearest third)	4.82*** (1.02)	2.20*** (0.67)	0.32 (0.30)
Breadth $\times$ middle third by distance	-0.54 (0.53)	0.02 (0.25)	0.09 (0.16)
Breadth $\times$ farthest third by distance	-0.77 (0.66)	0.24 (0.33)	0.07 (0.22)
Breadth $\times$ market outside the Northeast	-2.01*** (0.61)	-1.08** (0.47)	0.15 (0.21)
Mean claims per \$1,000	10.12	9.10	7.36
Market-years	18,618	30,096	69,139

Notes. Market cells of ordinary insurance, by period. Each column estimates equation (5.3) for claims per \$1,000 of insurance in force, interacting breadth with indicators for the insurer's headquarters state, its largest other market in the previous year, and the middle and farthest thirds of its other markets by distance from headquarters. The breadth coefficient is the differential in the nearest third; the interactions are differences from it. Panel B interacts breadth with local share and tenure instead. Panel C adds an interaction with markets outside the Northeast. Standard errors, clustered by firm and state-year, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table XI: Survival, Exit, and Surplus, by Size and Breadth
Panel A. Survival

Survival horizon	Breadth coefficient		Size (cond.)	<i>N</i>
	Unconditional	Conditional on size		
Five-year survival	0.052*** (0.008)	-0.001 (0.010)	0.104*** (0.014)	10,329
Ten-year survival	0.077*** (0.012)	0.001 (0.016)	0.152*** (0.020)	10,329

Panel B. Exit

Size group at exit (% of exits)	Consolidation	Failure	Withdrawal	<i>N</i>
Bottom third	82	12	5	129
Middle third	93	5	2	129
Top third	91	9	0	128
All exits	89	9	2	386

Panel C. Surplus and distress, New York statement firm-years

	Breadth	Conditional on size		MDE (breadth)	Mean	Firms
		Breadth	Size			
Surplus / liabilities	-0.175*** (0.064)	0.109 (0.096)	-0.529** (0.207)	0.268	0.202	72
S.d. of surplus / liabilities, 5 years	-0.070** (0.030)	0.040 (0.027)	-0.246*** (0.087)	0.077	0.046	59
Pr(surplus fell)	-0.051*** (0.017)	-0.017 (0.026)	-0.057 (0.037)	0.074	0.322	72
Pr(dividends cut by more than 10%)	-0.055* (0.029)	-0.009 (0.036)	-0.048 (0.057)	0.101	0.140	61

Notes. Ordinary-insurance firm-years, 1888–1940. Each cell is the coefficient on breadth, alone and conditional on a cubic in size, with region-by-year fixed effects; the size column reports the linear size term. Panel A: survival for five or ten more years. Panel B: exit channels. Panel C: firm-years matched to New York statements; surplus is relative to total liabilities, and *dividends cut* indicates dividends below 90 percent of the previous year's. MDE is the minimum detectable effect at 80 percent power. Standard errors are clustered by firm. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table XII: Pooling across Lives and across Places

<i>Panel A. Window variance components on breadth and size</i>					
	Breadth alone	Breadth and size		Breadth and size cubic	
	Breadth	Breadth	Size	Breadth	Size
Log variance of the loss ratio, $\ln R$	-0.588*** (0.051)	0.067 (0.054)	-0.949*** (0.052)	0.078 (0.056)	-1.020*** (0.081)
Log claim-count (lives) noise, $\ln N$	-1.253*** (0.081)	0.147*** (0.066)	-2.031*** (0.095)	0.140** (0.061)	-1.984*** (0.085)
Log variance net of lives noise, $\ln(R - N)$	-0.526*** (0.054)	0.025 (0.065)	-0.841*** (0.065)	0.055 (0.068)	-0.967*** (0.100)
Lives-noise share, N/R	-0.169*** (0.013)	0.008 (0.016)	-0.257*** (0.017)	0.011 (0.016)	-0.284*** (0.022)
Covariance share, $1 - B/R$	-0.048 (0.067)	-0.115 (0.086)	0.111 (0.097)	-0.080 (0.087)	-0.099 (0.118)
<i>Panel B. Medians by terciles of size and breadth</i>					
Size	Breadth	Windows	s.d. of loss ratio	Lives-noise share	Covariance share
Small	Narrow	1,603	0.086	0.63	0.04
	Middle	753	0.086	0.53	0.29
	Broad	73	0.073	0.72	0.16
Middle	Narrow	724	0.053	0.38	0.04
	Middle	1,210	0.058	0.27	0.22
	Broad	494	0.056	0.26	0.29
Large	Narrow	102	0.039	0.15	-0.01
	Middle	465	0.037	0.15	0.26
	Broad	1,861	0.034	0.08	0.42
				$\ln(R - N)$	$\ln(R - N) - 2 \ln \bar{L}$
<i>Panel C. Volatility relative to the level of losses: breadth coefficient (size cubic, region-by-year effects)</i>					
All windows				0.055 (0.068)	-0.223*** (0.063)
Windows starting 1888–1916				0.126 (0.097)	-0.387*** (0.094)
Windows starting 1921–1936				-0.009 (0.094)	-0.162* (0.083)
Non-overlapping windows				0.086 (0.083)	-0.184* (0.075)
Size coefficient, all windows				-0.967*** (0.100)	-1.053*** (0.094)
<i>Panel D. Components</i>					
$\ln$ mean loss ratio of the window				0.144*** (0.029)	
$\ln R$ and $\ln R - 2 \ln$ mean loss ratio				0.078 (0.056)	-0.209*** (0.052)
$\ln(R - N)$ controlling for $\ln$ mean loss ratio				-0.188*** (0.060)	
elasticity of $R - N$ to the mean loss ratio				1.70*** (0.11)	
<i>Panel E. Proportional geographic benchmark</i>					
$\ln G_r$, variance of proportional state shocks				-0.727*** (0.058)	
Share of G_r in $(R - N)/\bar{L}^2$, median [mean]				0.042 [0.156]	
Benchmark at the median share, and p that the estimate is not below it					-0.022 ($p = 0.001$)
Benchmark at the mean share, and p that the estimate is not below it					-0.084 ($p = 0.014$)
Windows (Panel C, row 1)				6,056	6,056

Notes. Five-year windows of consecutive firm-years, 1888–1940. R is the variance of the insurer’s loss ratio, N its claim-count component, losses $\times F/\text{premium}^2$ with F the average policy size, and B the variance it would have if its state loss ratios were uncorrelated. Panel A regresses each window statistic on breadth and size with region-by-year fixed effects. Panel B reports medians by terciles of size and breadth. Panels C–E use windows with $R > N$; $\bar{L}$ is the window’s mean loss ratio, and the relative outcome is $\ln(R - N) - 2 \ln \bar{L}$. Panel E compares the estimate with the reduction implied by removing G_r , the geographic variance G measured relative to the level of losses. Section 3.3 defines the measures. Standard errors are clustered by firm. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table XIII: The 1918–1919 Influenza Pandemic

<i>Panel A. Breadth and the dispersion of exposure (excess deaths per 1,000)</i>		
	Standard deviation	
Lowest breadth tercile	0.708	
Middle breadth tercile	0.639	
Highest breadth tercile	0.275	
Absolute deviation from the industry mean, on breadth	-0.170*** (0.043)	
<i>Panel B. Effect of excess deaths per 1,000 on the change from 1915–17 to 1918–19</i>		
	Loss ratio	Claims per \$1,000
Markets, within insurer	0.018* (0.010)	0.423 (0.379)
States, all insurers	0.030** (0.015)	0.835 (0.584)
Insurers, exposure from 1915–17 premium	0.011 (0.037)	2.078*** (0.654)
Insurers, exposure from 1912–14 premium	-0.008 (0.035)	1.262* (0.676)
Insurers, excluding the five highest-mortality states	0.006 (0.073)	3.205** (1.291)
<i>Panel C. Placebo and magnitudes</i>		
Placebo: change from 1912–14 to 1915–17 on exposure	0.025 (0.030)	0.903 (0.671)
R^2 of the insurer-level change on exposure	0.004	0.025
Standard deviation of the insurer-level change	0.125	3.998
Implied protection from breadth (lowest minus highest tercile)	0.0047	0.8987

Notes. Exposure is the premium-weighted average of 1918 excess deaths per 1,000 across the states in which an insurer wrote in 1915–1917. Panel A: dispersion of exposure by breadth tercile. Panel B: effect of excess deaths per 1,000 on the change in the loss ratio and in claims per \$1,000 from 1915–1917 to 1918–1919, within insurers across markets (row 1), across states (row 2), and across insurers (rows 3–5). Panel C: a placebo for 1912–1914 to 1915–1917 and the implied protection from breadth. Section 7.2 describes the design. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Online Appendix

Contents

A	Data Appendix	61
A.1	Source	61
A.2	Transcription and checks	61
A.3	Firms, exits, and headquarters	62
A.4	Coverage	62
A.5	Content of the state-level entries	62
A.6	Validity screen	63
A.7	Sample	64
A.8	Validation against the New York statements	65
A.9	Actual and expected mortality	65
A.10	The insurers' own accounts in the Spectator	66
A.11	Distance	68
B	Additional Results	69
B.1	Robustness of the cost of breadth	69
B.2	The insurers' own accounts	71
B.3	Investment income, expenses, and dividends	73
B.4	Actual versus expected mortality	73
B.5	Expansion and prior performance	74
B.6	The 1918 influenza pandemic	75
C	Online Appendix Tables and Figures	76

A Data Appendix

A.1 Source

The state panel is transcribed from the life-insurance volumes of the Spectator Company’s *Insurance Year Book*. Each volume reports, for every legal-reserve life insurer, the premiums it received and the losses it paid in each state and territory in the preceding calendar year, together with new insurance written and insurance in force, separately for ordinary, industrial, and group business where the company reported them. The state pages number approximately 10 in the volumes of the late 1880s and more than 50 by the late 1920s. Volumes are available for every year from 1888 through 1940. Fraternal benefit societies, which held a substantial share of the life insurance in force in the early twentieth century, are not reported in these pages and are therefore excluded from the panel.

A.2 Transcription and checks

The scanned state pages were transcribed by a text-recognition model that assigns each entry a confidence score. Every entry scored below 0.5 was re-keyed by hand, as were entries that failed two checks: the sum of firms’ entries in a state against the state totals printed in the volume, and each firm-state series against its own previous and following years. Firm names were then matched to a master list of companies, which assigns each insurer a single identifier. Remaining anomalies are addressed in the construction of the panel through indicator flags, which are retained rather than silently discarded:

- a firm-year whose total premium falls to near zero between two large neighboring years, or rises to five times the level of both neighboring years, is set to missing;
- a firm-year whose number of states with positive premium falls to 40 percent or less of both neighboring years is flagged as incompletely transcribed and excluded (318 firm-state-years), and a spuriously high count is flagged (478);
- a single-year gap in insurance in force is filled from the stock identity in $\text{force}_t = \rho \text{ in force}_{t-1} + \text{written}_t$, with the persistence ρ solved from the two neighboring years (832 cells), or by linear interpolation when that procedure fails (267 cells), and the imputed cells are flagged;
- firms observed in only a single year before 1940 are excluded.

A.3 Firms, exits, and headquarters

Each insurer carries one identifier across years. The *Year Book* reports each company’s home office and state of incorporation, and, for companies that ceased to write business, the year and manner of their departure: reinsurance of all of their policies by another company, merger, acquisition, failure, or voluntary retirement. Exits are recorded through 1950, so that ten-year survival is observed for every firm-year through 1940. A block of business assumed through reinsurance appears in the assuming company’s state pages as in-force business with little new insurance written; for this reason, the entry definition in Section 3.3 requires that the insurance in force of an entered market consist mostly of newly written business.

A.4 Coverage

The *Year Book* includes a page for every state and the District of Columbia from 1889, except Oklahoma (from 1893), Hawaii (1903), and Alaska (1920). The expansion of coverage over the panel period occurs in the set of companies itemized: 53 insurers in 1890, 72 in 1900, 206 in 1910, 257 in 1920, and about 320 in the 1930s (Online Appendix Table IA.6a, Panel B). The additional insurers were predominantly small, and the median number of states per insurer fell from 16 in 1890 to 3 in 1910. A company that first appears in the *Year Book* late in the panel appears with all the states in which it then operates; its first-year states are therefore treated as its founding markets rather than as entries; the entry definition also requires an entered market’s book to be mostly new business. Three quarters of insurers are covered continuously from their first to their last appearance, without a gap year or a flagged coverage dip or spike. The formation results are unchanged when the sample is restricted to them, to firms that operated in at most three states when they first appear, and to entries made after a state’s first two years in the *Year Book* (Table IA.6a).

A.5 Content of the state-level entries

Premiums are those received in the state and losses those paid in the state. The volumes do not specify whether a death claim is charged to the state in which the policy was sold or the state in which the claim was paid, and the two allocations differ when a policyholder relocates. The New York statements do not report business by state, with one exception: each company’s exhibit of its business in New York. For the 71 insurers observed in both sources, the Spectator’s New York cells match those exhibits closely. The correlation of logs is 0.93 for premium, 0.94 for losses paid, and

0.97 for insurance in force, and the median ratios of the Spectator figure to the statement figure are 1.06, 1.05, and 1.06 (Table [IA.2a](#), Panel B). The losses a company reported to the Spectator for a state are therefore, at least for New York, the losses it reported to that state’s regulator on its business there.

Two potential misallocations could affect the within-firm comparisons. First, if claims followed policyholders who moved, westward migration would move claims from the older, nearer markets that wrote the policies to newer, more distant ones, raising measured loss ratios in distant and recently entered markets; the evidence in the paper points in the opposite direction (Figure [IV](#), Panels C and D, and Table [VI](#)). Second, a company whose home office paid claims it could not attribute to a state might have charged them to its home state, which would raise its home-state loss ratio relative to its other markets. For companies headquartered in New York the comparison above covers the home state, and it provides no evidence of such a practice: the median ratio of Spectator to statement losses on New York business is 1.04 for the 28 companies domiciled in New York and 1.05 for the 36 domiciled elsewhere. Such a practice would, in any case, raise a broad insurer’s measured claims in its home state, which would work against the smaller home-state penalty reported in Section [5.4](#).

The losses column consists chiefly of death claims. Matured endowments are paid on the same schedule as death claims but constitute a return of saving rather than a mortality cost. Across 1,958 firm-years of insurers that filed in New York, the ratio of the sum of a firm’s Spectator ordinary losses to the death claims in its statement has a median of 1.00 and lies between 0.98 and 1.06 in every decade, whereas its ratio to death claims plus matured endowments is 0.88. Maturities are therefore only partly included, more so after 1920: the elasticity of Spectator losses relative to death claims with respect to the ratio of maturities to death claims is 0.22 to 0.29 before 1920 and 0.40 to 0.54 after, depending on the fixed effects (Table [IA.2a](#), Panel A), which suggests that some companies included maturities in reported losses, particularly after 1920.

A.6 Validity screen

A small number of cells report insurance in force that cannot be reconciled with the premiums or claims reported alongside it. I flag a cell when premium per \$1,000 in force exceeds 200 or falls below 5 (below 1 for group business, which is written as renewable term), claims per \$1,000 in force exceed 200, or the loss ratio exceeds 10. Among ordinary firm-state-years, the rule flags 1.9 percent of cells and 0.9 percent of premium: 720 cells with yield above 200, 708 below 5, 1,070 with claims

above 200, and 742 with a loss ratio above 10 (a cell can meet more than one condition). Flagged cells are more frequent in a market’s first year (18 percent of flagged cells, compared with 6 percent of all cells), when insurance in force is small and measurement error has the largest effect, and only 1 percent of them have imputed insurance in force. Cells with claims above 200 per \$1,000 have normal yields and a median loss of 46 percent of their insurance in force; their median year is 1930. For the New York cells that can be matched to the statements, all seven flagged cells differ from the statement by more than a factor of three in at least one column, compared with 9 percent of unflagged cells. Table [IA.3a](#) reports the main results under tighter and looser thresholds, under individual components of the rule, with cells winsorized rather than excluded, and without imputed or entry-year cells.

Two considerations indicate that the flagged cells are errors rather than outcomes. Annual death claims of 20 percent of the insurance in force are far outside the range consistent with the mortality of the period, and the yield bounds identify cells whose insurance in force is misreported relative to their premium. In six of the seven flagged New York cells, moreover, the two columns other than the discrepant one agree with the company’s statement. The claims-based flags are concentrated after 1920, when the losses reported by some companies appear to include matured endowments (Section [A.5](#)). Before 1920 the claims cost is similar with and without the screen, 18.7 and 16.6 percent of the mean, although a smaller share of it is attributable to the level component without the screen, 13.7 against 21.8 percent (Table [IA.3b](#)); over the full period the attribution depends on how cells with very high claims are treated, because those cells are concentrated after 1920.

A.7 Sample

Table [IA.1a](#) traces the samples used in the paper from the transcribed panel. Individual tables apply further outcome-specific restrictions (positive insurance in force for yields and claims per \$1,000, trimming of outcome tails, lags and leads, windows), which are stated in their notes.

The mean loss ratio differs across the paper’s tables, from about 0.24 to 0.38, because the unit (market or firm-year), the weighting, and the sample differ: firm-year means weight the many small insurers heavily, market means weight small markets, and the New York statement sample consists of large, older companies whose policies in force had been in force longer. The loss ratio of the New York statements themselves, defined from their accounting totals, averages 0.41 (Table [I](#), Panel C). Each table reports the mean of its own outcome.

A.8 Validation against the New York statements

For the 75 insurers that also filed with the New York Superintendent, the sum of a firm's state premiums can be compared with the premium income in its annual statement (Figure [IA.1](#)). Across about 2,000 firm-years the two series have a log-log slope of 1.04 and an R^2 of 0.93, and their correlation is between 0.94 and 0.98 in every decade. The median ratio of the state sum to statement premium is 0.93, consistent with statement premium including some income not allocated to states, and 65 percent of firm-years lie within 10 percent of that median ratio and 83 percent within 25 percent. Excluding the firm-years whose ratio lies more than 25 percent from the median leaves the breadth coefficients from regressions of the log loss ratio, log claims, and log yield on breadth and a cubic in size essentially unchanged: over the full period, for firm-years with positive losses and with the top and bottom percentile of each outcome trimmed, they are 0.128, 0.096, and -0.032 , compared with 0.119, 0.090, and -0.029 on all firm-years.

A.9 Actual and expected mortality

The ratio of actual to expected mortality in Table [IA.20a](#) is transcribed for 1909–1920 from Table XI of Part II of the *Annual Report of the Superintendent of Insurance of the State of New York* for 1910 through 1921, which reports for each life insurer licensed in New York the sources of the increase and decrease in its surplus during the preceding year, beginning with the expected mortality, the actual mortality, and the gain from mortality of its gain-and-loss exhibit. The figures are company-wide, include industrial business, and are computed on the net amount at risk under each company's valuation table. Each report's table is in three parts: New York companies, companies of other states, and, from 1916 to 1918, the U.S. branches of two German companies, which are outside the state panel. The reports for 1907–1909 do not contain such a table; for 1906–1908 the figures are transcribed from the mortality section of the gain-and-loss exhibit in each company's statement in Part II of those reports: the expected mortality on the net amount at risk, the actual mortality on the net amount at risk, and the gain or loss from mortality. These are the items Table XI later tabulates, and for 1909, the year in which both sources are available, the two coincide. Where a company files separate exhibits for its participating and non-participating, or ordinary and industrial, business and no combined exhibit, the departments are summed; for the two companies that report both, the departmental exhibits sum exactly to the combined exhibit. The statements provide 146 exhibits and 125 company-years for 1906–1908, and the panel for 1906–1920 has 551

company-years of 58 insurers, of which 54 are in the state panel.

For every exhibit of 1906–1908, expected less actual mortality equals the gain from mortality to the cent; in four exhibits, one of the three lines is illegible and is derived from the other two, and no printed totals are available for verification in those years. Every life insurer licensed in New York in the statement data for 1906–1908 is covered. For 1909–1920, expected less actual mortality equals the gain from mortality in every company-year, and in each report the columns sum to the printed totals of each part and the parts to the printed aggregate, with the following exceptions. In 1911 the other-state companies’ actual mortality sums to \$39,000 more than the printed total and their gain to \$39,000 less (0.07 percent of actual mortality), while every row satisfies the identity; the rows are retained as printed. In 1916 one printed expected mortality is inconsistent with its row and with the column total and is corrected, and in 1918 one partially illegible figure is resolved by the same two checks. Two 1918 rows deviate from the identity by \$1. In every year the companies transcribed are exactly the life insurers licensed in New York in the statement data, with the exception of a reinsurer whose row is blank. The actual mortality of Bankers Life of Iowa, which began as an assessment company, is footnoted in 1911–1913 as including amounts applied in reduction of assessments; its A/E is about 0.97 through 1914 and 0.65 thereafter, so its firm-years before 1915 are excluded.

A.10 The insurers’ own accounts in the Spectator

The company-level data of Table [IA.13](#) are transcribed from five parts of the life volumes of the *Insurance Year Book*, each of which reports the business of two years earlier. The “Life Insurance History” of the 1905 and 1912 volumes gives, for each company, a row for every year since its organization: premiums, interest, death claims, matured endowments, surrender values, dividends, expenses, new business and insurance outstanding by number and amount, assets, liabilities, and surplus (2,939 company-years of 1886–1910). The compendium exhibits of the 1894–1909 and 1912 volumes rank the companies by new business issued and placed, by insurance terminated by mode, and by whole-life, endowment, and other insurance in force, and from the 1907 volume an exhibit of expenses reports medical fees and inspections separately. From the 1913 volume the compendium prints the full annual statement of every company, which I transcribe for the volumes not digitized earlier (3,160 statements); the volumes of 1905, 1910, 1915, 1920, and 1925 had been digitized before and are converted to the same items. The gain-and-loss exhibit, printed from the 1897 volume and absent from the 1898 volume and from the missing volume of 1911, gives each company’s expected and actual mortality (3,290 company columns, for 1895 and 1897–1926).

The pages were transcribed from the scanned images, with the scans' text layer as a first draft and every figure verified against the image. Each exhibit was checked against its accounting identities (total income equals premiums plus other income, terminations by mode sum to total terminations, insurance in force by type sums to the total, and expected less actual mortality equals the gain from mortality); figures that fail an identity after re-reading are printed misprints and are kept as printed. The same firm-year transcribed from two different tables agrees closely: insurance in force in the History and in the compendium agree within 2 percent in 95 percent of 1,148 firm-years, and death claims in the History and in the earlier digitizations in 97 percent of 228. Each printed company name was linked to the identifier of the state panel by its name and home city and confirmed by the agreement of its premiums and insurance in force with the firm's state sums; uncertain links were checked by hand, and links that remain uncertain are not used. The median ratio of company death claims to the sum of the firm's state-table losses is 1.00. For the 463 firm-years of 42 insurers that also appear in Table [IA.20a](#), the Spectator's A/E has a correlation of 0.989 with the ratio from the New York reports, and expected mortality agrees within 1 percent in 94 percent of them.

The company accounts of combination companies mix their ordinary and industrial departments, whereas the state panel is ordinary business only, so firm-years in which the insurer wrote industrial business are excluded from the company measures of Table [IA.13](#), as are firm-years whose company insurance in force differs from the state sum by more than a factor of 1.5. Insurance in force is measured as the average of the current and previous year-ends where both are reported. The gain-and-loss exhibits of the period are on each company's own valuation table and basis, so A/E compares the mortality of the insurer's policies in force with the table the insurer chose.

The valuation basis is transcribed from the 1924 edition of *Reports of Life Insurance Companies*, which prints for each legal-reserve company a description of the basis on which its policy reserves were valued, by table, interest rate, and the years of issue to which each applied, together with its form of organization and its ratio of actual to expected mortality for 1922. I record whether any business was valued on the Actuaries' (Combined Experience) table and on the American Experience table, the lowest and highest valuation interest rates, and whether the company used preliminary-term or full level-premium valuation. For business valued on more than one basis the variables record each basis that applied. Companies were linked to the panel by name and home state and confirmed by the agreement of the 1922 ratio with the Spectator's gain-and-loss exhibit for that year: 288 companies are linked, 272 of them with a complete basis. Of these, 42 percent

valued some business on the Actuaries' table, 96 percent on the American Experience table, and 78 percent used preliminary-term valuation.

A.11 Distance

Distance is the great-circle distance from the city of a firm's home office to the geographic centroid of the state. For large states, the centroid may lie far from the locations where business was written, so I recompute distance to each state's largest city. The two measures have a correlation of 0.94 in logs, and the within-firm gradients of Table VI are similar: -0.625 per log-kilometer for premium yield, compared with -0.739 , and -0.033 for the loss ratio, compared with -0.039 , all significant at the 1 percent level (Table IA.26a, Panel C). The *Year Book* does not report the locations of general agencies consistently enough to measure distance to them.

B Additional Results

Tables and figures in this appendix use the same screened sample as the main tables unless otherwise stated.

B.1 Robustness of the cost of breadth

This section collects the robustness checks on Table VIII referred to in Section 5.2.

Age, tenure, and logarithms. Controlling for firm age, the premium-weighted tenure of the firm’s markets, and an indicator for firms already writing in 1888 leaves the full-period estimates at 11.6, 8.6, and -3.9 percent, and in logarithms, on the firm-years with positive losses, they are 12.2, 9.0, and -3.2 .

Sensitivity to outliers and small claim counts. The firm-year aggregates of narrow insurers are noisier than those of broad insurers because they are based on fewer claims, so a percentile trim of the aggregates disproportionately removes narrow firm-years: over the full period the joint trim drops 6.3 percent of single-state firm-years and 0.6 percent of the broadest third of the rest. Table IA.7b re-estimates Table VIII without the trim, with the trim applied separately within four breadth groups, and in a Poisson pseudo-maximum likelihood model of claims with insurance in force as the exposure, which retains every firm-year and treats the zero or few claims of a market with few insured lives as small counts. Before 1920 the claims cost is 13.4 percent without the trim (standard error 4.6), 15.2 with the trim applied within breadth groups, and 37 log points in the Poisson model, and its level component is 16.4, 18.4, and 13 log points; the Poisson level component is estimated from a cell-level Poisson model with firm-year and state-year effects. Over the full period the claims cost is less robust: without the trim it is 4.7 percent with a standard error of 3.8, although it is 8.3 percent with the trim within breadth groups and 15 log points in the Poisson model. The loss-ratio costs are statistically significant in every version, and the pricing costs in every version except the full-period Poisson aggregate (-2.6 , standard error 2.7). I therefore treat the pre-1920 estimates as the paper’s main result and the full-period estimates as secondary.

Precision of the level term. The level term of a firm-year is estimated from the markets in which the firm writes in that year, and for a single-state insurer it equals that market’s residual (Section 5.1). Excluding single-state firm-years raises the pre-1920 claims level component to 23.5

percent. Under Poisson claim counts, the variance of a firm-year’s level term is inversely proportional to the number of lives it insures, regardless of their distribution across states; weighting firm-years by their lives in force, and hence by the precision of their level terms, raises the pre-1920 claims cost to 48.1 percent and its level component to 54.7 (Table [IA.7b](#)). The difference between broad and narrow insurers in the Spectator’s losses is larger, not smaller, among the insurers whose level terms are most precisely estimated. These weighted estimates are upper bounds on the difference in death claims paid: weighting gives most of the weight to the largest insurers, whose state-table losses exceed their own death claims in proportion to breadth, and weighted by premium the difference in own death claims is 7.0 percent in 1906–1920 (Section [B.2](#) and Table [IA.2b](#)).

Nonlinearity in breadth. The log specification imposes a monotone, concave relation. Replacing breadth with indicators for bins of N_{eff} shows that the claims cost increases steadily with breadth up to 20 effective states: over the full period, relative to firms with at most 1.25 effective states, from 5 percent at 1.25 to 2 effective states to 32 percent at 10 to 20, and before 1920 from 1 to 54 percent (Table [IA.7c](#)). The largest insurers of the period, New York Life, Metropolitan, Prudential, the Equitable, and the Mutual (before 1920 Northwestern Mutual rather than Prudential ranks among the five), wrote in 8 to 15 effective states and therefore fall in the bins where the cost is largest; excluding them leaves the estimates unchanged, at 19.4 percent before 1920 and 8.9 over the full period. Above 20 effective states claims are lower, but that bin contains only 36 firm-years, 35 of them after 1924, of five small companies; three of them, Massachusetts Protective, Paul Revere, and United Benefit, belonged to accident-and-health insurance groups and wrote small ordinary insurance in force dispersed across many states. Excluding those firm-years raises the full-period claims cost to 11.4 percent.

The extensive margin of the distance gradient. Distant positions are small, with 0.75 log points less insurance in force per log-kilometer, and once the market’s insurance in force is held fixed the no-claims gradient of Table [VI](#) disappears (on the common sample of Table [IA.9a](#), from 8.2 percentage points to -0.5), while the gradient of the loss ratio among markets with positive claims steepens from -0.016 to -0.030 and the Poisson gradient of losses on premium from -0.042 to -0.072 .

Further specifications. Table [IA.12a](#) addresses alternative explanations that would generate the gradients without a difference between firms. Under firm fixed effects the yield dilution is larger,

−2.83 against −1.01, so it does not reflect a time-invariant difference across firms (Panel A). On the New York sample, controls for the composition of assets and for surplus reduce the yield dilution from −0.98 to −0.75, which is no longer statistically significant, while the loss-ratio coefficient is similar. By line of business (Panel B), the loss-ratio cost is specific to ordinary business, which was individually underwritten and locally priced. The yield dilution also holds for industrial business, where it is larger in absolute terms, −6.1 per \$1,000, because industrial policies carried premiums of about 44 per \$1,000 of coverage and substantially higher collection costs; relative to that mean it is about 14 percent. Within markets (Panel C), adding log distance raises the claims coefficient from 0.770 to 0.981, and headquarters-region-by-year effects leave it at 0.851.

B.2 The insurers’ own accounts

This section reports the evidence from the insurers’ own company accounts and New York statements summarized in Section 5.2. The claims differential appears in the death claims the insurers themselves paid. The Spectator printed each company’s accounts, which extend the comparison to every insurer (Table IA.13, Panel A). On firm-years whose company insurance in force agrees with the state tables within a factor of 1.5, a standard deviation of breadth raised insurers’ own death claims per \$1,000 of insurance in force by 15.4, 17.5, and 13.8 percent of the mean in 1888–1905, 1906–1920, and 1921–1926, against 23.3, 18.8, and 14.7 percent in the state-table claims of the same firm-years. The estimate before 1906 is imprecise and is attributable mainly to the insurers licensed in New York, whose own statements show the difference directly: 21.0 percent of the mean before 1906 (standard error 8.0; wild-cluster bootstrap $p = 0.035$ with 41 insurers) and 4.4 percent (13.6) after (Table IA.14, Panel A). The difference lies in how often insured lives died. A standard deviation of breadth raised deaths per 1,000 policies in force by 16 to 17 percent in every period, whereas the average claim relative to the average policy was larger only before 1906 (Table IA.13, Panel B).

This evidence describes the typical insurer rather than the typical premium dollar. Weighted by premium, the difference in own death claims is 12.8, 7.0, and 6.1 percent of the weighted mean, none statistically significant, whereas the state-table claims of the same firm-years remain large: among the largest insurers the Spectator’s losses exceeded their own death claims in proportion to breadth. Weighted by premium, broad insurers still recorded more deaths per policy in every period, although only marginally so in 1906–1920. Estimates of the claims differential that weight firm-years by their size are therefore upper bounds on the difference in the death claims the largest

insurers paid.¹⁸

The accounts also allow an examination of whether the differential reflects the type of policies broad insurers sold rather than the lives they selected. A broad insurer before 1920 was typically an older company, and an older portfolio of policies insures older lives; claims per \$1,000 also differ across whole-life, endowment, and term policies and with the size of policies. Policy mix does not explain the differential. Holding size fixed, the endowment share of insurance in force does not vary with breadth in the company accounts, and controlling for it and for the average policy size leaves the difference in own death claims at 14.3, 15.9, and 13.3 percent of the mean in the three periods (Table [IA.13](#)). The age of the policies in force is more difficult to exclude. Among the insurers that report a gain-and-loss exhibit in 1906–1920, broad insurers’ tabular mortality per dollar of insurance in force was 4.4 log points higher (standard error 1.3), about a quarter of the difference in their actual mortality (Table [IA.13](#), Panel C), so their policies in force were slightly older. Firm age is measured from the first year of the panel, however, and for the companies founded before 1888, which include several of the broadest, the age of the policies in force is not observed, so differences in duration beyond those that tabular mortality records cannot be excluded. The remaining difference is chiefly a difference in the mortality of the lives broad insurers accepted.¹⁹

Participating business does not account for the differential either. Before 1906 the largest companies competed for new business through deferred-dividend policies sold by costly agency forces (Section [2.1](#)). In 1888–1906, when the eight deferred-dividend writers collected 59 percent of premium, breadth raised claims by 18.6 percent of the mean including them and 16.6 percent excluding them, and the differential was as large in 1907–1920, after the Armstrong legislation ended new deferred-dividend business. Holding size fixed, broad insurers paid no higher policyholder dividends per dollar of premium, and controlling for the dividend rate does not reduce the level term (Table [IA.14](#), Panel C).²⁰

¹⁸Table [IA.2b](#) reports the ratio of state-table losses to own death claims, with and without matured endowments and annuities; Table [IA.7b](#) the weighted estimates; and Table [IA.8a](#), Panel D, the estimates with lapses as a control, which raise the difference to 21.1, 19.7, and 16.4 percent.

¹⁹On the New York firm-years of 1888–1905 broad insurers’ reserves per dollar of insurance in force were, if anything, lower, and controlling for reserves and the endowment share leaves the difference in their own death claims at 19.2 percent against 21.0 (Table [IA.14](#), Panel B). Controls for policy mix and recent writings on the New York firm-years, and for firm age and market tenure in Table [VIII](#), Panel D, do not reduce the claims differential appreciably (Table [IA.2a](#)).

²⁰Table [IA.18a](#) lists the deferred-dividend writers (the Equitable, the Mutual, New York Life, and five insurers whose deferred-dividend liability was at least 5 percent of liabilities) and reports the estimates: 16.7 percent in 1907–1920, and an interaction of breadth with an indicator for these insurers of 16.7 with a standard error of 18.1, too imprecise to exclude a difference. With organizational form as a proxy for managerial discipline (Jensen, 1986; Berger and Ofek, 1995), the differential does not differ between stock companies and mutuals (Section [B.5](#)).

B.3 Investment income, expenses, and dividends

Investment income could offset a higher loss ratio if broad insurers earned higher returns on their reserves. The evidence indicates that they did not (Table [IA.2a](#)). Conditional on size, breadth is unrelated to the yield an insurer earned on its assets and to its gain from operations per dollar of premium, and broad insurers paid somewhat higher commissions per dollar of premium, 1.9 points per standard deviation. Policyholder dividends and expenses per dollar of premium do not vary with breadth, and the cost does not differ between mutual and stock companies (Table [IA.18b](#)), so lower yields do not reflect participating business priced gross of dividends.

B.4 Actual versus expected mortality

Among the insurers licensed in New York, relative to the mortality their valuation tables predicted for the ages and policies of their insureds, broad insurers' mortality was, if anything, lower rather than higher, and controlling for it does not reduce the claims difference on those firm-years; across all insurers the Spectator's exhibits show the opposite (Table [IA.13](#), Panel C). The gain-and-loss exhibit that the reforms of 1906 added to the annual statements reports, for every life insurer licensed in New York, the mortality its insureds experienced during the year and the mortality its valuation table predicted for them.²¹ The ratio of the two, A/E, is the industry's own measure of selection: it holds the age and policy composition of an insurer's policies in force fixed at the rates the table assigns to it, although not the time since the lives were selected, since valuation tables are ultimate tables. I transcribe it for 1906–1920, 551 company-years of 58 insurers, and match 521 firm-years of 48 insurers to the preferred sample; these are the insurers licensed in New York, which include the largest of the period.²² The mean A/E is 0.73, and 0.98 in the influenza year 1918.

Table [IA.20a](#), Panel A, reports the estimates. A standard deviation of breadth lowers A/E by 0.046 without the size control and by 0.085 with it. With region-by-year effects, the preferred specification, it lowers A/E by 0.063 with a standard error of 0.045, 9 percent of the mean and not statistically significant, and when firm-years are weighted by expected mortality, which gives the largest insurers the most weight, by 0.124 (0.051). Excluding the three industrial insurers, the

²¹For 1906–1908 the figures are taken from the mortality section of each company's exhibit in Part II of the Superintendent's reports for 1907–1909; from 1909 the Superintendent tabulated them in Table XI of Part II, which sets out the sources of the increase and decrease in each company's surplus. For 1909, when both sources exist, the two agree. The figures are company-wide, include industrial business, and are computed on the net amount at risk. Section [A.9](#) describes the transcription and its checks.

²²With region-by-year effects, firm-years that are the only observation in their region-year drop out, leaving 517 firm-years of 46 insurers in Table [IA.20a](#), Panel A, column 3.

estimate is -0.032 (0.045). By period it is 0.008 (0.063) in 1906–1908 and -0.064 (0.041) in 1909–1920. In every year from 1906 to 1920 the insurers above the median breadth had a lower aggregate A/E than those below it, 0.70 against 0.90 in 1906 and 0.92 against 1.19 in 1918 (Figure [IA.7](#)). With 46 insurers, a wild-cluster bootstrap gives $p = 0.19$ for the preferred estimate and $p = 0.029$ for the weighted one (Table [IA.21a](#)). The 95 percent confidence interval of the preferred estimate excludes an increase in A/E of more than about 3 percent of its mean. If less favorable selection accounted for the claims difference, A/E would have had to rise with breadth by approximately as much as claims per dollar of coverage do, 16 to 19 percent of the mean. The years 1906–1908 alone are too few to exclude such an increase.

Nor does A/E account for the claims difference (Panel B). On the same firm-years, breadth raises the claims level component by 16.0 percent of the mean, imprecisely estimated on 46 insurers. A/E itself predicts claims, by 11.2 per \$1,000 per unit of A/E, as expected; controlling for it raises the breadth coefficient to 21.1 percent, and the aggregate and loss-ratio estimates move in the same direction. In 1906–1908 the claims level component on these firm-years is 45.3 percent of the mean (standard error 16.6), and 44.9 with A/E as a control (Panel C). A broad insurer incurred higher claims per dollar of coverage even though the mortality of its insureds, relative to the rates its table assigned to their ages and policies, was no higher than that of a narrow insurer’s insureds.

The A/E evidence covers 1906–1920 and the insurers licensed in New York, and for those insurers the difference in their own death claims after 1906 is small and imprecisely estimated (Table [IA.14](#), Panel A). In the New York reports A/E is therefore available only where the difference it would have to explain may be small. The Spectator printed the same exhibit for nearly every insurer from 1906 (Section [A.10](#)). On the firm-years of Table [IA.20a](#) that also have a Spectator exhibit, the two sources give the same estimate, -0.011 (0.041) from the New York reports and -0.008 (0.041) from the Spectator, and the estimate of -0.063 on all of Table [IA.20a](#) comes from the firm-years the Spectator lacks, chiefly 1909, for which the Spectator printed no exhibit (Table [IA.15a](#), Panel D). Across all insurers, A/E rises with breadth, by 0.053 (0.014) in 1906–1920 and 0.046 (0.010) in 1921–1926 (Section [5.4](#) and Table [IA.13](#)).

B.5 Expansion and prior performance

If a deterioration in the quality of the insured lives preceded expansion, insurers whose loss ratios deteriorated should have been more likely to enter new states. Table [IA.16a](#) estimates, over the

firm-year panel,

$$\text{Entry}_{i,t+1} = \sum_{j=0}^J \pi_j LR_{i,t-j} + x'_{it}\beta + \alpha_i + \alpha_t + \varepsilon_{it}, \quad (\text{B.1})$$

where $\text{Entry}_{i,t+1}$ equals 100 if firm i opens at least one new state in year $t + 1$, LR_{it} is its loss ratio in markets entered more than two years earlier, x_{it} collects size, age, breadth, new-business intensity, yield, premium growth, and the firm's share of its home-state market, and α_i and α_t are firm and year effects. All regressors are standardized, and the unconditional entry rate is about 15 percent.

Across insurers, a standard deviation of the loss ratio is followed by 1.96 fewer percentage points of entry. Within an insurer the relation disappears: under firm fixed effects the contemporaneous term is 0.41 with a standard error of 0.68, and with three lags the four terms sum to -0.06 ($p = 0.40$). Entry was instead preceded by growth: new-business intensity raises the probability of entry by 2.7 to 5.4 percentage points per standard deviation, and it is the only characteristic that raises entry under firm effects. Entry declined with age and, within an insurer, with breadth itself, by 10.7 points per standard deviation. Insurers whose existing markets were deteriorating did not generally respond by entering new states, and claims in the markets an insurer already held did not rise in the years before an expansion episode (Section 5.3).

The agency account of expansion predicts a larger cost where managers were subject to less discipline. In this period the contrast most often drawn was between stock companies and mutuals. On the insurers licensed in New York, whose statements identify mutual companies, the breadth coefficients do not differ between them (Table IA.18b).

B.6 The 1918 influenza pandemic

Section 7 reports the analysis of the 1918–1919 influenza pandemic. Table IA.25b reports the power of the insurer-level test in the specification of Table IA.26a, Panel E, which uses all of an insurer's markets and region effects. Restricting the sample to insurers with small claim-count variance, whose loss ratios are less affected by the number of claims, leaves too few insurers to improve the precision of the estimate.

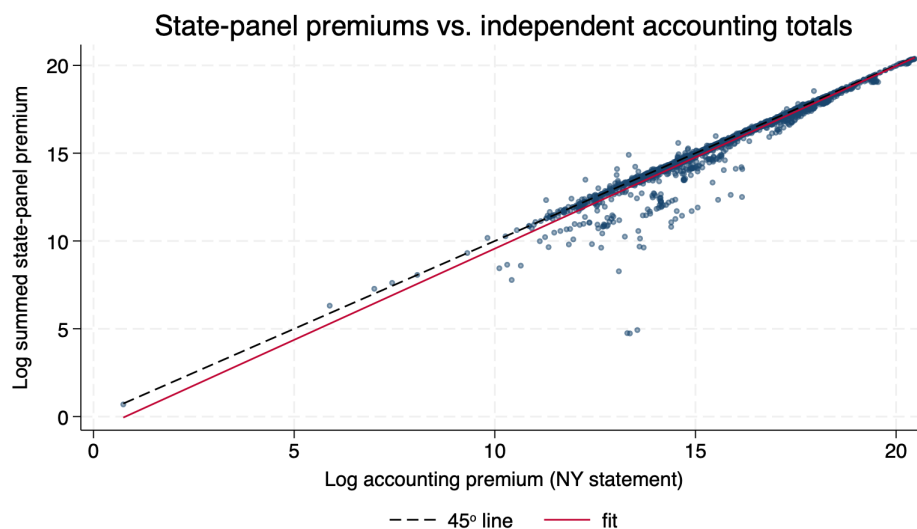
C Online Appendix Tables and Figures

Table IA.1: Sample construction and *Year Book* coverage
IA.1a. Sample construction

	Firm-state- years	Firm- years	Insurers	Used in
Transcribed state cells, all lines, 1888–1940	161,125	13,019	905	
Insurers domiciled in the United States	156,002	11,082	819	
Excluding incompletely transcribed firm-years	155,684	11,056	819	
Positive premium	151,928	10,962	818	
States and the District of Columbia	149,581	10,952	818	Section 3
Passing the validity screen, all lines	146,650	10,831	817	Table I
Ordinary business, passing the screen	123,870	10,341	781	Analysis sample
with positive insurance in force	121,498	10,221	780	Table VIII, before trim
after the joint trim of Table VIII		9,897	768	Table VIII, Panels C–D
of which 1888–1920		4,162	416	Table VIII, Panel A
of which 1921–1940		5,733	611	Table VIII, Panel B

Notes. Table I, Panel B, reports 10,287 firm-years of the screened all-lines sample: it excludes the firm-years with no recorded losses, and 10,081 have a loss ratio after its percentile trim. The regressions of Table VIII drop firm-years that are alone in their region-year cell, so their reported N is slightly smaller (9,897 over 1888–1940). Other samples: market entries followed for ten years, 1,426 (Figure V; 1,562 measured from the entry year); five-year windows, 7,285 of 521 insurers (Table XII); firm-years matched to a New York statement, about 1,860 (Table XI, Panel C). Periods are standardized separately, and the 1888–1920 and 1921–1940 firm-years are trimmed within their periods; the subperiod samples therefore do not sum to the full-period sample.

Figure IA.1: Aggregated state-panel premiums and New York statement premium income



Sources: author's calculations from the Spectator *Insurance Year Book* state pages and the New York Superintendent's annual-statement accounting data. Each point is a firm-year for a firm appearing in both sources. The horizontal axis is the log of the firm's accounting income-statement premium; the vertical axis is the log of the firm-year total of its state-panel premiums (placed in dollars). The dashed line is the 45-degree line and the solid line is the ordinary-least-squares fit.

Table IA.2: What the state-table losses contain
IA.2a. Checks against the New York statements

<i>Panel A. Spectator ordinary losses against New York statement terminations (firm-years)</i>			Elasticity to $\log(1 + \text{maturities/deaths})$		
	Median ratio of losses to				
	death claims	deaths + maturities	1888–1940	1888–1920	1921–1940
Median ratio, all firm-years	1.00	0.88			
elasticity, year FE			0.285** (0.117)	0.218* (0.124)	0.543*** (0.207)
elasticity, firm and year FE			0.332*** (0.118)	0.291** (0.132)	0.404*** (0.133)
<i>Panel B. Spectator New York cells against the New York statement's exhibit of business in New York</i>					
	Corr. of logs	Median ratio	Within 10%	Within 25%	
Premium	0.93	1.06	0.59	0.86	
Losses paid	0.94	1.05	0.30	0.63	
Insurance in force	0.97	1.06	0.63	0.90	
Share of cells off by more than 3×: flagged / unflagged	1.00	0.09			
<i>Panel C. Cost of breadth, New York firm-years (% of the outcome mean)</i>					
	Baseline	+ product mix	+ seasoning	+ maturities	All
Claims per \$1,000	7.9 (7.7)	9.4 (7.1)	9.7 (6.6)	7.8 (7.7)	9.9* (6.0)
Loss ratio	7.4 (9.0)	8.8 (7.7)	9.2 (8.1)	7.4 (9.1)	9.6 (6.5)
Premium per \$1,000	-1.6 (1.8)	-2.0 (1.5)	-1.6 (1.9)	-1.7 (1.8)	-2.1 (1.5)
<i>Panel D. Breadth, investment income and distress, New York firm-years</i>					
	Breadth (uncond.)	Breadth (size cubic)	Size (size cubic)		
Investment yield on assets	0.001** (0.001)	-0.000 (0.001)	0.003*** (0.001)		
Gain from operations / premium	0.025 (0.031)	0.018 (0.045)	0.006 (0.057)		
Gain from operations / assets	-0.010 (0.007)	-0.004 (0.011)	-0.013 (0.013)		
Commissions / premium	0.006 (0.008)	0.019** (0.009)	-0.016 (0.013)		
Surplus / liabilities	-0.175*** (0.064)	0.109 (0.096)	-0.529** (0.207)		
Surplus fell	-0.051*** (0.017)	-0.017 (0.026)	-0.057 (0.037)		
Policyholder dividends cut by 10%+	-0.055* (0.029)	-0.009 (0.036)	-0.048 (0.057)		
Surrender values / premium	0.015 (0.011)	-0.009 (0.016)	0.045** (0.021)		

Panel A compares the sum of a firm's Spectator ordinary losses with the death claims and matured endowments in its New York statement; the elasticity is the coefficient of the log ratio of Spectator losses to death claims on the log of one plus the ratio of maturities to death claims, for firm-years with a ratio between 0.2 and 5. Panel B compares Spectator New York cells with the statement's exhibit of business in New York; a cell is classified as differing by more than a factor of three if premium, losses, or insurance in force differs by that much, among cells with all six figures positive. Panels C and D use New York firm-years in the sample of Table VIII, Panel C. Product mix: endowment and term shares of insurance in force and the log average policy; seasoning: writings in the current and four previous years over insurance in force; maturities: maturities over deaths plus maturities. Investment yield is interest and dividends over mean assets; gain from operations is total revenue less total expenditure. Region-by-year effects, a cubic in size where indicated, and firm clusters.

Table IA.2: What the state-table losses contain (continued)
IA.2b. State-table losses and the insurers' own payments

	1888–1905	1906–1920	1921–1926
<i>Panel A. Unweighted (log points per s.d. of breadth)</i>			
Losses / death claims, unweighted	1.3 (4.6)	-3.5** (1.7)	-2.6 (1.8)
same, firm-years reporting maturities	4.1 (6.9)	-2.2 (3.3)	-2.4 (2.2)
Losses / (death claims + matured endowments), unweighted	7.8 (11.3)	0.8 (4.0)	-3.6* (2.1)
Losses / (death claims + endowments + annuities), unweighted	7.8 (11.3)	0.8 (3.9)	-3.8* (2.1)
ln(1 + maturities / death claims), unweighted	-6.1 (8.6)	-2.5 (3.5)	1.0 (1.9)
<i>Panel B. Weighted by premium (log points per s.d. of breadth)</i>			
Losses / death claims, premium-weighted	19.9** (10.0)	16.2** (6.8)	16.3** (6.6)
same, firm-years reporting maturities	21.2* (10.9)	22.7*** (7.9)	18.2** (7.2)
Losses / (death claims + matured endowments), premium-weighted	24.5** (10.2)	28.8*** (9.7)	17.3** (8.1)
Losses / (death claims + endowments + annuities), premium-weighted	24.5** (10.2)	28.5*** (9.7)	16.9** (8.3)
ln(1 + maturities / death claims), premium-weighted	-4.8 (11.9)	-5.8 (7.2)	1.0 (3.9)
Firm-years (Panel A, row 1)	587	2,202	886
Firm-years with maturities reported (row 2)	426	898	529

Firm-years of Table IA.13, Panel A, row 1, with positive own death claims and state-table losses. Each outcome is the log of the Spectator's state-table losses over the insurer's own payments, times 100: death claims; death claims plus matured endowments (matured endowments paid where the company statement or the *History* reports them, else the face amount terminated by maturity); and death claims, matured endowments, and payments to annuitants. The last row of each panel is ln(1 + maturities/death claims), times 100. Rows 2–5 keep firm-years with maturities reported. The log ratios are trimmed at the 1st and 99th percentiles within each period. Panel B weights firm-years by premium. Size cubic, region-by-year effects, firm clusters. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table IA.3: The validity screen
IA.3a. Robustness to the treatment of implausible cells

Screen	Cells dropped (%)	Claims per \$1,000		Premium per \$1,000		Loss ratio	Covariance share	
		Total	Level	Total	Level		Narrowest	Broadest
Baseline: yield 5–200, claims ≤ 200 , loss ratio ≤ 10	1.9	9.6*** (3.0)	10.3*** (3.2)	-3.5*** (0.8)	-2.9*** (0.8)	12.2*** (3.0)	0.09	0.44
Tighter: yield 10–150, claims ≤ 100 , loss ratio ≤ 5	3.7	11.2*** (2.9)	12.9*** (3.2)	-2.8*** (0.8)	-2.2*** (0.8)	13.0*** (3.0)	0.10	0.49
Looser: yield 2–400, claims ≤ 400 , loss ratio ≤ 20	0.8	6.3* (3.3)	6.5* (3.8)	-5.1*** (1.0)	-4.7*** (1.0)	9.5*** (3.5)	0.08	0.40
Loosest: yield 1–1000, claims ≤ 1000 , loss ratio ≤ 50	0.4	4.9 (3.6)	0.1 (4.3)	-5.8*** (1.1)	-5.7*** (1.2)	8.5** (3.9)	0.09	0.36
Yield bounds only	1.1	7.4** (3.8)	-0.3 (4.0)	-3.4*** (0.8)	-2.8*** (0.8)	9.3** (3.9)	0.10	0.44
Claims and loss-ratio bounds only	1.0	8.2*** (2.9)	8.2** (3.4)	-6.7*** (1.2)	-8.1*** (1.2)	13.2*** (3.1)	0.08	0.38
No screen	0.0	6.6* (3.6)	-3.0 (3.9)	-7.2*** (1.2)	-10.9*** (1.9)	11.5*** (3.8)	0.09	0.35
No screen, cells winsorized at 1/99	0.0	2.3 (5.1)	1.3 (5.3)	-9.3*** (1.3)	-9.3*** (1.3)	11.5** (5.0)	0.09	0.40
No screen, cells winsorized at 5/95	0.0	6.5** (2.9)	7.1** (3.0)	-3.9*** (0.8)	-3.6*** (0.7)	9.9*** (3.0)	0.09	0.47
Baseline, imputed in-force cells dropped	2.6	9.6*** (3.0)	10.3*** (3.2)	-3.5*** (0.8)	-2.8*** (0.8)	12.1*** (3.0)	0.09	0.47
Baseline, entry-year cells dropped	7.6	10.3*** (2.9)	10.7*** (3.2)	-2.8*** (0.8)	-2.3*** (0.8)	12.3*** (3.0)	0.10	0.45
1888–1920 only, baseline screen	1.4	18.5*** (4.4)	21.7*** (4.3)	-5.6*** (1.3)	-4.7*** (1.3)	22.4*** (4.5)	0.13	0.34
1888–1920 only, no screen	0.0	16.6*** (4.4)	13.7*** (4.8)	-9.6*** (1.9)	-10.2*** (2.0)	23.9*** (4.6)	0.12	0.01

Each row re-estimates the specification of Table VIII, Panel C, and the pooling statistics of Section 6 under the indicated treatment of market cells. The row for 1888–1920 with the baseline screen is a separate run of Table VIII, Panel A; it retains one fewer firm-year at the trim boundary than Table VIII and differs from it by up to 0.2 points. *Total* is the coefficient on breadth for the firm aggregate and *level* the coefficient on the firm-year term of the decomposition, both in percent of the outcome mean, conditional on a cubic in size. The covariance share is the median $1 - B/R$ in the narrowest and broadest breadth quartiles. The winsorized rows clip cell premium and claims per \$1,000 at the stated percentiles, recover insurance in force and losses from the clipped values, and omit the firm-year percentile trims, which are ill-defined when many cells are clipped to the same value. “Imputed” cells have insurance in force filled from neighboring years; “entry-year” cells are the first year of a market entered after 1888. Standard errors are clustered by firm. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table IA.3: The validity screen (continued)
 IA.3b. The exact decomposition without the validity screen, 1888–1920

	Firm aggregate β^Y	Level β^ϕ	Markets β^α	Position β^δ	N
<i>Panel A. With the validity screen (Table VIII, Panel A; % of the outcome mean)</i>					
Loss ratio	22.6*** (4.5)	25.8*** (4.5)	-0.9 (0.9)	-2.3*** (0.9)	4,162
Claims per \$1,000 in force	18.7*** (4.4)	21.8*** (4.3)	-1.9 (1.2)	-1.2 (1.1)	4,162
Premium per \$1,000 in force	-5.6*** (1.3)	-4.7*** (1.3)	-0.8** (0.4)	-0.1 (0.2)	4,162
<i>Panel B. Without the validity screen</i>					
Loss ratio	23.9*** (4.6)	25.9*** (5.0)	-1.6 (1.8)	-0.5 (0.9)	4,194
Claims per \$1,000 in force	16.6*** (4.4)	13.7*** (4.8)	0.7 (2.2)	2.1** (1.0)	4,194
Premium per \$1,000 in force	-9.6*** (1.9)	-10.2*** (2.0)	0.5 (0.6)	0.1 (0.3)	4,194

Panel A repeats Table VIII, Panel A. Panel B re-estimates it on firm-years built from all ordinary market cells with positive insurance in force, without the validity screen of Section A.6; the joint trim, the decomposition, and the controls are as in Table VIII. Each cell reports the coefficient on breadth (standardized log effective number of states) in percent of the outcome mean, conditional on a cubic in size, with region-by-year effects; the level, markets, and position columns sum exactly to the firm aggregate. Standard errors, clustered by firm, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table IA.4: Market structure and the persistence of leaders
IA.4a. Rank persistence and mobility

Panel A. Rank persistence: Spearman correlation of national premium (survivors)

Horizon	5 years	10 years	20 years
Spearman correlation	0.91	0.86	0.80

Panel B. Ten-year size-quintile transition matrix (row %)

	Exit	Q1	Q2	Q3	Q4	Q5 (top)
Q1	50.5	12.7	15.7	12.3	7.4	1.5
Q2	40.5	2.8	19.0	23.9	11.3	2.5
Q3	24.9	1.6	5.7	31.4	29.2	7.3
Q4	14.6	1.1	1.0	6.9	47.4	29.0
Q5 (top)	5.6	0.2	0.2	0.5	4.7	88.8

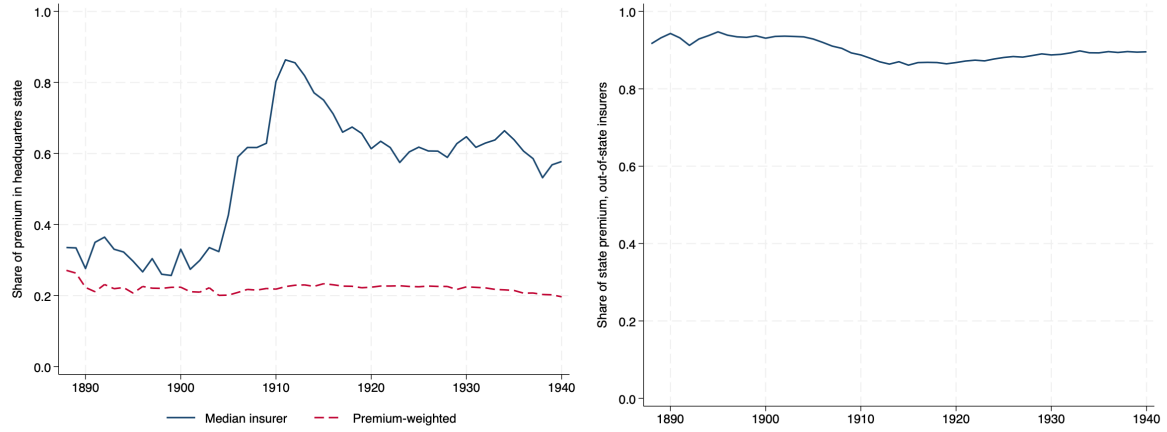
Notes. The ordering of firms by national premium is highly stable. Panel A reports the Spearman correlation of national premium income between year t and year $t+h$ across surviving firms. Panel B reports the ten-year transition matrix across within-year size quintiles (Q1 smallest, Q5 largest), with an absorbing “Exit” column for firms no longer observed after ten years. Base years are restricted to those with an observable $t+10$ outcome, and firms alive but temporarily unobserved at $t+10$ are dropped rather than coded as exits.

Table IA.5: Market entries by decade and region

Decade	Northeast	Midwest	South	West	All	Per 100 firm-years
1890s	8	44	92	38	182	32.5
1900s	32	127	222	135	516	46.4
1910s	25	203	232	154	614	25.7
1920s	67	282	416	193	958	34.0
1930s	41	93	158	102	394	13.0

Number of market entries, as defined in Section 3.3, by decade and census region of the entered state, and entries per 100 firm-years of the insurers active in the decade. Figure II plots the counts.

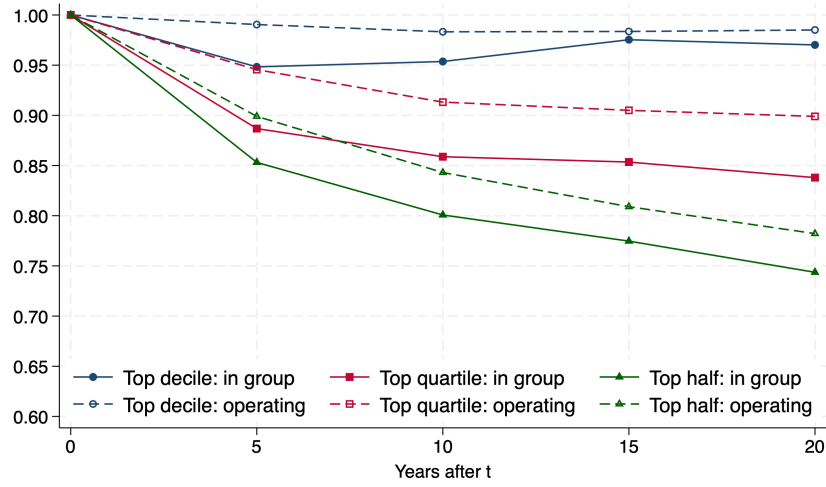
Figure IA.2: The geographic structure of the industry, annual series



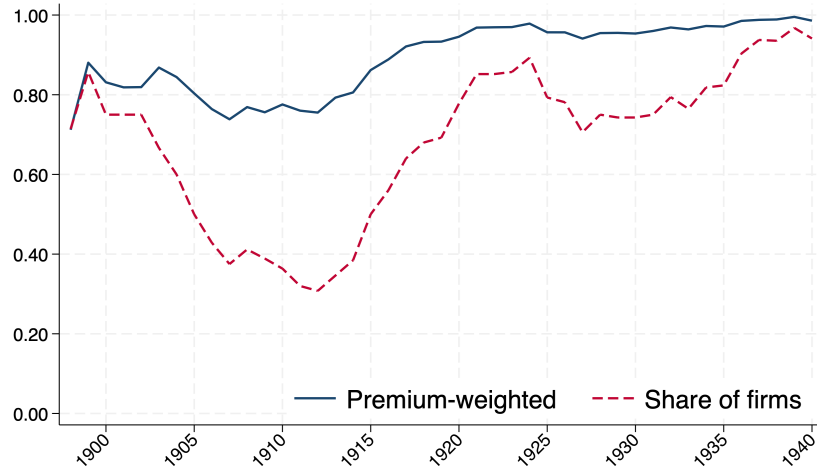
Panel A: the share of an insurer's premium collected in its headquarters state, median across insurers and premium-weighted. Panel B: the share of a state's premium written by insurers headquartered in other states, averaged across states. Table III reports selected years.

Figure IA.3: Persistence of top firms

Panel A. Retention by top-group definition



Panel B. Incumbent share



Notes. Panel A plots retention curves for firms in the top premium decile in year t : the share still in the top decile and the share still operating, as a function of the horizon h . Panel B plots the incumbent share of the top decile over time, where an incumbent is a current top-decile firm that was also in the top decile ten years earlier; the share is reported both premium-weighted and as a share of firms. Base years for Panel B are restricted to those with a full ten-year lookback.

Table IA.6: Entry payoffs, predictability, and the ordering of breadth and size
IA.6a. Entry payoffs, concentration, and coverage

<i>Panel A. Entry payoffs and concentration</i>								
	Payoff	Entries	Aban-	Share	Median	Top-decile share of		
	base		doned	> 10×	survivor multiple	year-10 premium	payoff multiple	within state size terciles
All entries	Second year	1,426	0.21	0.06	1.9	0.49	0.56	0.48
	Entry year	1,562	0.19	0.22	4.1	0.49	0.74	0.48
Continuously covered firms	Second year	927	0.17	0.05	1.8	0.47	0.53	0.46
	Entry year	1,019	0.16	0.22	3.9	0.48	0.71	0.47
After a state's first two years	Second year	1,422	0.21	0.06	1.8	0.49	0.56	0.48
	Entry year	1,557	0.19	0.22	4.1	0.49	0.74	0.48
At most three states when first seen	Second year	978	0.24	0.05	1.9	0.49	0.55	0.48
	Entry year	1,060	0.22	0.22	4.4	0.50	0.74	0.49
<i>Panel B. Year Book coverage (screened ordinary cells)</i>								
Year		States	Insurers	Median states per insurer				
1890		48	53	16.0				
1900		49	72	11.5				
1910		50	206	3.0				
1920		51	257	4.0				
1930		51	318	6.0				
1940		51	321	4.0				

Panel A: entries as in Figure V. The payoff multiple is the market's premium in years nine to eleven after entry relative to its premium in the entry year or in the second year, divided by the firm's own premium multiple over the same window; abandoned entries count as zero under either base. Top-decile shares are the share of the total accounted for by the best 10 percent of entries, ranked by year-ten premium, by the payoff multiple, or by year-ten premium within terciles of the entered state's total premium in the entry year (average of the three terciles). Continuously covered firms appear in every year from their first to their last appearance, without a flagged coverage dip or spike. Panel B: screened ordinary cells.

Table IA.6: Entry payoffs, predictability, and the ordering of breadth and size (continued)
IA.6b. Breadth and size

	+5 years		+10 years	
	Size	Breadth	Size	Breadth
Breadth (t)	0.095*** (0.015)	0.909*** (0.015)	0.107*** (0.026)	0.839*** (0.027)
Size (t)	0.739*** (0.022)	0.007 (0.017)	0.665*** (0.038)	0.011 (0.031)
Observations	6911	6911	4940	4940
R2	0.83	0.86	0.77	0.79

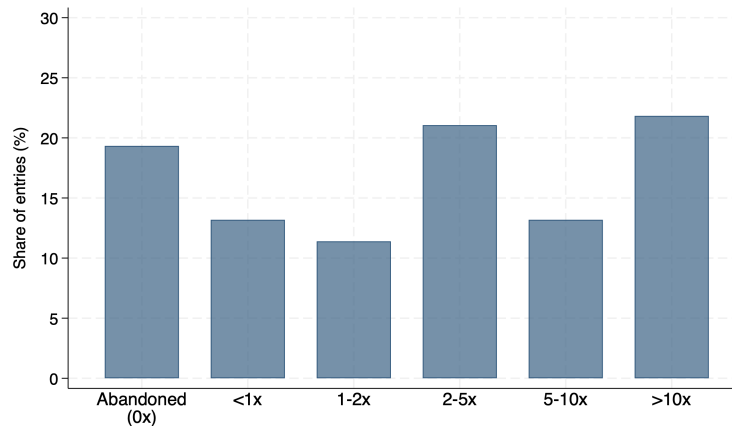
Notes. Size is the standardized log of firm premium; breadth is the standardized log of the effective number of states ($N_{\text{eff}} = 1/\text{HHI}$). The sample is ordinary-insurance firm-years, 1888–1940. Each column regresses one variable's value h years ahead on the current values of both variables, so every column includes the lagged dependent variable and the cross term; columns 1 and 3 are the size equation, columns 2 and 4 the breadth equation, at horizons of five and ten years, with region $\times$ year fixed effects and standard errors clustered by firm. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table IA.6: Entry payoffs, predictability, and the ordering of breadth and size (continued)
 IA.6c. Formation: entry payoffs, predictability, and the ordering of breadth and size

<i>Panel A. Ten-year entry payoffs (Figure V) under alternative bases and samples</i>						
	Entries	Abandoned	< 1×	> 10×	Median survivor multiple	Top-decile premium share
Entry-year base	1,486	0.18	0.13	0.22	4.07	0.48
Second-year base	1,292	0.16	0.22	0.06	1.86	0.47
Mean of years 1–2 as base	1,292	0.16	0.17	0.08	2.35	0.47
Firms first observed 1890+, entry-year base	1,276	0.20	0.13	0.22	4.15	0.49
Firms first observed 1890+, second-year base	1,114	0.17	0.22	0.06	1.88	0.48
Including entries of firms that exit (payoff 0)	2,164	0.44	0.09	0.15	4.07	0.58
<i>Panel B. Can winners be identified? (5-fold out-of-sample, log payoff, survivors)</i>						
	Entries	OOS R^2	Top-decile precision			
Adding first-year premium and early loss ratio	1,057	0.314	0.43			
Pre-entry information only	1,057	0.050	0.11			
<i>Panel C. Abandonment within ten years on early performance (per s.d., year FE)</i>						
Relative premium growth, years 1–2	-0.021** (0.010)					
Loss ratio, years 1–3	0.016 (0.012)					
<i>Panel D. Table IA.6b: breadth → size and size → breadth</i>						
	h	Breadth→Size	Size→Breadth	N		
Full sample (paper)	5	0.095*** (0.015)	0.007 (0.017)	6,911		
Full sample (paper)	10	0.107*** (0.026)	0.011 (0.031)	4,940		
Firms first observed 1890+	5	0.112*** (0.015)	-0.010 (0.021)	5,085		
Firms first observed 1890+	10	0.142*** (0.025)	-0.017 (0.038)	3,347		
Incumbents of 1888–89	5	-0.034 (0.029)	0.082** (0.037)	1,785		
Incumbents of 1888–89	10	-0.078 (0.050)	0.115 (0.077)	1,557		
Size as 3-year average	5	0.058*** (0.016)	0.043** (0.018)	5,758		
Size as 3-year average	10	0.041 (0.026)	0.066** (0.032)	4,151		
Firm and year FE	5	0.061** (0.026)	0.054** (0.024)	6,882		
Firm and year FE	10	0.015 (0.027)	0.081** (0.035)	4,921		

Panel A follows the construction of Figure V, which uses the second-year base, but sets the first sample year before the markets held in 1888 are removed, so its entries differ from those of Figure V and Table IA.6a, Panel A (1,292 against 1,426 on the second-year base, 1,486 against 1,562 on the entry-year base). Panel B predicts the log ten-year payoff (second-year base) of surviving entries with pre-entry information (distance, firm premium, firm age, the market's premium, five-year growth, and number of insurers, and the entry year) or adding the entry's first-year premium and early loss ratio; "top-decile precision" is the share of entries predicted to be in the top decile that actually fall in it. Panel C is a linear probability model with entry-year effects. Panel D repeats Table IA.6b.

Figure IA.4: Market entry payoffs measured from the entry year



As Figure V, Panel A, with the payoff multiple measured from the market's premium in its entry year.

Table IA.7: Specification of the firm-level cost: functional form, trims, estimator, and weights
IA.7a. Functional form of breadth: logs and levels

Breadth measure	Loss ratio		Claims per \$1,000		Premium per \$1,000	
	% of mean	Within R^2	% of mean	Within R^2	% of mean	Within R^2
Standardized log N^{eff} (preferred)	12.2*** (3.0)	0.026	9.6*** (3.0)	0.033	-3.5*** (0.8)	0.022
Standardized N^{eff}	8.8*** (3.2)	0.022	7.1** (3.2)	0.031	-2.8*** (0.7)	0.019
Both: log N^{eff}	15.0*** (5.7)	0.026	11.1** (5.5)	0.033	-3.3** (1.5)	0.022
N^{eff}	-3.1 (5.6)		-1.6 (5.6)		-0.1 (1.1)	
<i>Step function (omitted: $N^{\text{eff}} \leq 1.25$)</i>						
N^{eff} 1.25–2	5.7 (3.8)		5.0 (3.9)		-0.6 (1.4)	
N^{eff} 2–3.5	11.1** (4.6)		7.3 (4.6)		-4.4*** (1.6)	
N^{eff} 3.5–6	20.4*** (6.3)		15.3*** (5.8)		-4.9*** (1.8)	
N^{eff} 6–10	30.8*** (8.0)		24.5*** (7.7)		-7.7*** (2.2)	
N^{eff} 10–20	37.5*** (8.8)		31.6*** (8.6)		-8.9*** (2.4)	
N^{eff} above 20	-30.9* (16.2)		-39.2*** (14.2)		-15.3*** (4.0)	

Firm-year sample of Table VIII, Panel C. Coefficients in percent of the outcome mean, conditional on a cubic in size, with region-by-year effects and firm clusters. The step function replaces breadth with indicators for bins of the effective number of states, $N_{\text{eff}} \leq 1.25$ omitted; the bin above 20 contains 36 firm-years (Table IA.7c).

Table IA.7: Specification of the firm-level cost: functional form, trims, estimator, and weights (continued)

IA.7b. The cost of breadth under alternative treatments of outlying firm-years, small counts, and the precision of the level term

	Claims per \$1,000		Loss ratio		Premium per \$1,000	
	Aggregate	Level	Aggregate	Level	Aggregate	Level
<i>Panel A. 1888–1920</i>						
Preferred specification (Table VIII)	18.7*** (4.4)	21.8*** (4.3)	22.6*** (4.5)	25.8*** (4.5)	-5.6*** (1.3)	-4.7*** (1.3)
No firm-year trim	13.4*** (4.6)	16.4*** (4.5)	17.9*** (4.9)	20.7*** (4.9)	-8.9*** (1.8)	-7.9*** (1.7)
Trim within breadth groups	15.2*** (4.6)	18.4*** (4.4)	18.8*** (4.5)	22.0*** (4.4)	-6.6*** (1.5)	-5.7*** (1.5)
Poisson, exposure offset (log points)	36.7*** (6.6)	12.9*** (4.0)	42.3*** (7.6)	19.7*** (4.3)	-6.3** (2.8)	-6.2*** (1.4)
level with book-size deciles (log points)		10.7*** (4.0)		18.4*** (4.3)		-7.2*** (1.4)
Without single-state firm-years	20.8*** (4.5)	23.5*** (4.5)	24.6*** (4.7)	27.1*** (4.8)	-5.3*** (1.4)	-4.4*** (1.4)
Weighted by lives in force	48.1*** (7.6)	54.7*** (7.8)	57.5*** (8.0)	63.6*** (8.3)	-10.6** (4.3)	-9.7** (4.2)
Weighted by insurance in force	32.0*** (5.4)	34.6*** (5.8)	36.1*** (5.8)	38.6*** (6.1)	-5.0** (2.0)	-4.5** (1.9)
Without the five largest insurers	19.4*** (4.8)	23.0*** (4.7)	22.5*** (4.9)	26.1*** (4.9)	-5.1*** (1.3)	-4.2*** (1.3)
Without firm-years above 20 effective states	18.6*** (4.4)	21.8*** (4.3)	22.6*** (4.5)	25.8*** (4.5)	-5.6*** (1.3)	-4.7*** (1.3)
<i>Panel B. 1888–1940</i>						
Preferred specification (Table VIII)	9.6*** (3.0)	10.3*** (3.2)	12.2*** (3.0)	12.9*** (3.5)	-3.5*** (0.8)	-2.9*** (0.8)
No firm-year trim	4.7 (3.8)	4.9 (3.9)	8.3** (3.9)	8.4** (4.1)	-7.1*** (1.3)	-6.5*** (1.2)
Trim within breadth groups	8.3*** (3.0)	9.2*** (3.2)	11.1*** (3.1)	12.2*** (3.6)	-4.9*** (1.0)	-4.3*** (1.0)
Poisson, exposure offset (log points)	14.7*** (4.8)	7.3** (2.9)	17.1*** (5.0)	12.4*** (2.9)	-2.6 (2.7)	-5.0*** (1.0)
level with book-size deciles (log points)		4.2 (2.8)		9.7*** (2.9)		-5.6*** (1.0)
Without single-state firm-years	10.0*** (2.9)	12.0*** (3.1)	12.6*** (3.0)	15.1*** (3.3)	-3.3*** (0.8)	-2.7*** (0.8)
Weighted by lives in force	23.2*** (3.7)	25.0*** (3.9)	25.7*** (3.8)	26.9*** (4.0)	-3.1* (1.8)	-2.5 (1.8)
Weighted by insurance in force	16.6*** (3.2)	16.6*** (3.2)	18.2*** (3.7)	18.2*** (3.7)	-1.5 (2.6)	-1.2 (2.7)
Without the five largest insurers	8.9*** (3.0)	10.0*** (3.3)	11.3*** (3.1)	12.3*** (3.6)	-3.3*** (0.8)	-2.7*** (0.8)
Without firm-years above 20 effective states	11.4*** (2.9)	12.2*** (3.1)	14.0*** (3.0)	14.8*** (3.4)	-3.4*** (0.9)	-2.8*** (0.8)

Each cell is the coefficient on breadth for the firm aggregate or for the level term of the exact decomposition, conditional on a cubic in size, with region-by-year effects and firm clusters; standard errors below. Rows other than the Poisson rows are in percent of the outcome mean. *Trim within breadth groups* applies the joint trim of Table VIII separately to single-state firm-years and to terciles of breadth among the rest. *Poisson*: the aggregate row is a Poisson pseudo-maximum-likelihood regression of firm-year losses (premium) with log insurance in force (log premium for the loss ratio) as exposure, on all firm-years without a trim; the level row estimates equation (5.1) by Poisson on market cells, with the same exposures, firm-year and state-year effects, and the position term, and regresses the estimated firm-year effect on breadth (firm-years whose markets record no losses have no finite effect and drop out); coefficients in log points. *Lives in force* weights firm-years by insurance in force divided by average policy size, the inverse of the variance of the level term under Poisson claim counts. The weighted rows rest on the state-table losses of the largest insurers and are upper bounds on the difference in death claims paid (Table IA.2b). The five largest insurers are those with the most ordinary premium in the period (for 1888–1920: New York Life, the Mutual, the Equitable, Northwestern Mutual, and Metropolitan; for 1888–1940, Prudential in place of Northwestern). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table IA.7: Specification of the firm-level cost: functional form, trims, estimator, and weights (continued)

IA.7c. The cost of breadth across the distribution of breadth

N_{eff}	1888–1920				1888–1940			
	Claims Aggregate	Level	Premium Aggregate	Firm- years	Claims Aggregate	Level	Premium Aggregate	Firm- years
<i>Panel A. All insurers</i>								
1.25–2	0.8 (4.6)	0.6 (4.5)	-3.9** (1.7)	663	5.0 (3.9)	0.2 (4.4)	-0.6 (1.4)	1,581
2–3.5	12.7* (6.9)	9.8 (6.6)	-6.3*** (2.3)	483	7.3 (4.6)	-0.8 (5.2)	-4.4*** (1.6)	1,416
3.5–6	12.0* (6.8)	14.0** (6.8)	-8.0*** (2.3)	447	15.3*** (5.8)	12.3* (6.3)	-4.9*** (1.8)	1,056
6–10	40.3*** (11.2)	43.2*** (10.4)	-12.8*** (2.9)	482	24.5*** (7.7)	22.4*** (7.8)	-7.7*** (2.2)	1,131
10–20	54.1*** (12.7)	65.2*** (12.6)	-14.9*** (3.6)	592	31.6*** (8.6)	36.9*** (9.2)	-8.9*** (2.4)	1,216
Above 20				1	-39.2*** (14.2)	-45.0*** (16.3)	-15.3*** (4.0)	36
<i>Panel B. Without the five largest insurers</i>								
1.25–2	0.0 (4.7)	-0.1 (4.6)	-3.9** (1.7)	663	5.0 (4.0)	0.0 (4.5)	-0.7 (1.4)	1,579
2–3.5	11.8* (7.1)	8.9 (6.8)	-6.4*** (2.3)	481	7.2 (4.7)	-1.2 (5.3)	-4.5*** (1.6)	1,413
3.5–6	11.1 (7.1)	13.5* (7.0)	-7.9*** (2.3)	443	14.7** (6.0)	11.8* (6.4)	-5.0*** (1.8)	1,049
6–10	44.5*** (11.8)	48.0*** (10.9)	-12.2*** (2.9)	435	26.0*** (7.8)	24.3*** (8.0)	-7.4*** (2.2)	1,006
10–20	56.2*** (14.3)	69.3*** (14.1)	-13.2*** (3.6)	491	27.8*** (9.2)	35.4*** (10.0)	-8.2*** (2.5)	1,099
Above 20				1	-38.9*** (14.0)	-45.8*** (16.5)	-15.6*** (4.0)	36

Firm-year sample of Table VIII for each period. Breadth is replaced by indicators for bins of the effective number of states ($N_{\text{eff}} \leq 1.25$ omitted), conditional on a cubic in size, with region-by-year effects and firm clusters; coefficients in percent of the outcome mean. *Level* is the level term of the exact decomposition. Bins with fewer than ten firm-years are not reported. The five largest insurers had 8 to 15 effective states. The 36 firm-years above 20 effective states belong to Massachusetts Protective (14, 1925–1940), United Benefit (10, 1931–1940), Paul Revere (9, 1932–1940), Union Mutual of Iowa (2), and Provident Savings (1, 1908). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table IA.8: Size measure, denominators, sample checks, and firm dynamics
IA.8a. Size measured by insurance in force, denominators, and further checks

	1888–1920	1921–1940	1888–1940
<i>Panel A. Table VIII with size measured by insurance in force (% of mean)</i>			
Claims per \$1,000, size = premium (Table VIII)	18.7*** (4.4)	3.8 (3.6)	9.6*** (3.0)
level term	21.8*** (4.3)	2.5 (3.9)	10.3*** (3.2)
Claims per \$1,000, size = insurance in force	20.8*** (4.5)	5.4 (3.6)	11.5*** (3.0)
level term	23.6*** (4.5)	4.3 (3.9)	12.5*** (3.2)
Loss ratio, size = premium (Table VIII)	22.6*** (4.5)	6.8* (3.7)	12.2*** (3.0)
level term	25.8*** (4.5)	5.6 (4.3)	12.9*** (3.5)
Loss ratio, size = insurance in force	19.4*** (4.6)	5.8 (3.7)	10.7*** (3.0)
level term	22.4*** (4.6)	5.0 (4.3)	11.8*** (3.5)
Premium per \$1,000, size = premium (Table VIII)	-5.6*** (1.3)	-3.1*** (1.0)	-3.5*** (0.8)
level term	-4.7*** (1.3)	-2.5*** (1.0)	-2.9*** (0.8)
Premium per \$1,000, size = insurance in force	-0.6 (1.2)	-0.8 (1.0)	-0.3 (0.8)
level term	0.2 (1.2)	-0.3 (0.9)	0.2 (0.8)
<i>Panel B. Denominators</i>			
Claims per \$1,000 of year-end in force, 1888–1920	18.9*** (4.1)		
of mean in force, same firm-years	17.4*** (3.9)		
	Aggregate	Level term	
<i>Panel C. The 1888–1920 claims cost: further checks (% of mean)</i>			
Clustered by headquarters state	18.7*** (5.3)	21.8*** (5.6)	
Without firm-years within five years of exit	18.1*** (4.6)	21.2*** (4.4)	
Insurers with at least 15 firm-years	22.9*** (5.8)	27.0*** (5.6)	
Without insurers writing industrial business	17.8*** (5.8)	19.3*** (5.4)	
	1888–1905	1906–1920	1921–1926
<i>Panel D. Company accounts, Table IA.13 sample</i>			
State-table over company mean insurance in force, $100 \times \log$	-7.7** (3.9)	-4.7** (2.3)	-3.6** (1.7)
Own death claims, firm-years with lapses reported (% of mean)	20.4** (9.1)	17.3*** (4.8)	13.3*** (3.6)
controlling for lapses per \$1,000	21.1** (8.9)	19.7*** (4.7)	16.4*** (3.6)
A/E on new business / insurance in force (per s.d.)	0.039 (0.034)	-0.025** (0.010)	-0.004 (0.010)
New business / insurance in force on breadth	-0.058 (0.045)	0.007 (0.011)	0.032** (0.013)

Panel A: Table VIII with the size cubic in standardized log premium (as in that table) or in standardized log insurance in force. Panel B: claims per \$1,000 of year-end insurance in force and of the mean of the current and previous year-end, on the 1888–1920 firm-years with consecutive years. Panel C: the 1888–1920 claims cost with standard errors clustered by headquarters state, without firm-years within five years of an exit, on insurers with at least fifteen firm-years, and without insurers that wrote industrial business in any year. Panel D: Table IA.13 firm-years; the log ratio of the state-table sum of insurance in force to the company's mean insurance in force, insurers without industrial business, trimmed at the 1st and 99th percentiles; own death claims with and without lapses per \$1,000 as a control; A/E on the standardized ratio of new business to mean insurance in force, with breadth and size; and that ratio on breadth. Size cubic, region-by-year effects, firm clusters unless noted. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table IA.8: Size measure, denominators, sample checks, and firm dynamics (continued)
IA.8b. The claims cost before 1920: firm dynamics, size, pricing, and magnitude

	Aggregate	Level term
<i>Panel A. Firm dynamics, 1888–1920 (claims per \$1,000, % of the aggregate mean)</i>		
Baseline (common sample)	19.3*** (4.5)	22.5*** (4.6)
+ size five years earlier (cubic)	14.7*** (4.5)	18.8*** (4.6)
+ growth over five years	16.8*** (4.3)	20.6*** (4.5)
+ decline from peak size	18.7*** (4.5)	21.9*** (4.6)
+ new business / insurance in force	21.8*** (4.2)	24.4*** (4.4)
+ firm age and 1888 incumbency	16.7*** (4.1)	21.5*** (4.5)
All of the above	16.7*** (4.0)	21.3*** (4.5)
Firm-years / insurers		2,423 / 260
<i>Panel B. Breadth by size quartile, Table VIII sample, claims aggregate</i>		
Smallest quarter	4.3 (10.3)	
Second	0.2 (6.6)	
Third	22.5*** (5.4)	
Largest quarter	34.6*** (8.2)	
<i>Panel C. Underwriting yield, 1888–1920, local-share quintiles in the position term (% of mean)</i>		
Yield aggregate	-5.57*** (1.27)	
level term	-1.92 (1.25)	
markets term	-1.29** (0.50)	
position term	-2.36*** (0.34)	
<i>Panel D. Magnitude, 1888–1920</i>		
Claims per \$1,000 per s.d. of breadth	1.49	
as % of premium per \$1,000	4.6	
Excess claims from breadth beyond size, % of losses	1.74	
% of premium	0.65	

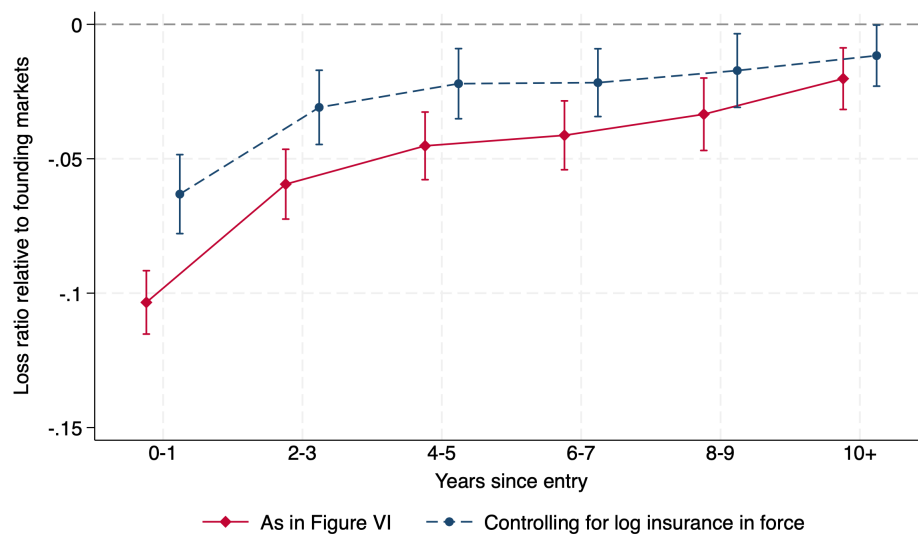
1888–1920 preferred sample of Table VIII. Panel A: firm-years with premium five years earlier observed; size five years earlier is the log of that premium, standardized as current size, in a cubic; growth is the annualized log change of premium over five years; the decline from peak is the log of the highest premium to date over current premium; new business is insurance written over insurance in force; growth, the decline, and new business are trimmed at the 1st and 99th percentiles. Firm-year premium, insurance in force, and insurance written are summed over the screened cells of all years. Panel B: breadth interacted with quartiles of size, with quartile indicators. Panel C: the exact decomposition of premium per \$1,000 with quintiles of the firm's share of the state's premium added to the position term, in percent of the mean. Panel D: the claims coefficient in dollars per \$1,000 and in percent of mean premium per \$1,000; the excess claims are the coefficient times the positive part of the residual of breadth on the size cubic and region-by-year effects, times insurance in force, summed over firm-years, in percent of total losses and premium. Region-by-year effects, firm clusters. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table IA.9: The position term: market size
IA.9a. Market size and the position term

<i>Panel A. Log distance within a firm-year (Table VI, Panel A)</i>						
	Loss ratio	Pr(no claims)	Loss ratio if claims > 0	Premium per \$1,000	Losses, Poisson	
As in Table VI	-0.0379*** (0.0037)	0.0815*** (0.0066)	-0.0157*** (0.0043)	-0.739*** (0.087)	-0.0418*** (0.0126)	
+ log insurance in force	-0.0231*** (0.0038)	-0.0052** (0.0026)	-0.0301*** (0.0037)	-0.581*** (0.086)	-0.0718*** (0.0151)	
+ in-force deciles	-0.0280*** (0.0036)	0.0137*** (0.0020)	-0.0257*** (0.0038)	-0.587*** (0.087)		
+ tenure	-0.0280*** (0.0033)	0.0535*** (0.0054)	-0.0145*** (0.0039)	-0.583*** (0.089)	-0.0434*** (0.0143)	
+ tenure and log insurance in force	-0.0197*** (0.0035)	-0.0032 (0.0025)	-0.0256*** (0.0035)	-0.506*** (0.090)	-0.0683*** (0.0155)	
Log insurance in force on log distance					-0.746*** (0.044)	
<i>Panel B. Loss ratio by years since entry (Figure IV, Panel C)</i>						
Years since entry	As in Figure IV	Common sample	+ log in force	+ in-force deciles	Poisson	Poisson + log in force
0–1	-0.106*** (0.006)	-0.103*** (0.006)	-0.063*** (0.007)	-0.052*** (0.007)	-0.263*** (0.089)	-0.373*** (0.089)
2–3	-0.060*** (0.007)	-0.059*** (0.007)	-0.031*** (0.007)	-0.034*** (0.007)	-0.129*** (0.042)	-0.214*** (0.039)
4–5	-0.045*** (0.006)	-0.045*** (0.006)	-0.022*** (0.007)	-0.030*** (0.007)	-0.087** (0.043)	-0.160*** (0.037)
6–7	-0.041*** (0.007)	-0.041*** (0.007)	-0.022*** (0.006)	-0.031*** (0.007)	-0.123** (0.061)	-0.189*** (0.056)
8–9	-0.034*** (0.007)	-0.033*** (0.007)	-0.017** (0.007)	-0.027*** (0.007)	-0.110** (0.051)	-0.170*** (0.049)
10+	-0.020*** (0.006)	-0.020*** (0.006)	-0.012** (0.006)	-0.019*** (0.006)	-0.038 (0.036)	-0.082*** (0.028)
<i>Panel C. Exact decomposition with in-force deciles in the position term (% of mean)</i>						
	1888–1920		1888–1940			
	Position	+ in-force deciles	Position	+ in-force deciles		
Claims: aggregate	18.7*** (4.4)	18.7*** (4.4)	9.6*** (3.0)	9.6*** (3.0)		
level	21.8*** (4.3)	19.6*** (4.2)	10.3*** (3.2)	8.4*** (3.1)		
position	-1.2 (1.1)	1.6 (1.2)	-1.6 (1.5)	0.9 (1.5)		
Loss ratio: aggregate	22.6*** (4.5)	22.6*** (4.5)	12.2*** (3.0)	12.2*** (3.0)		
level	25.8*** (4.5)	24.9*** (4.4)	12.9*** (3.5)	11.8*** (3.5)		
position	-2.3*** (0.9)	-1.1 (1.0)	-1.9 (1.9)	-0.5 (1.9)		

Panel A repeats Table VI, Panel A, on the cells with positive insurance in force, adding the log of the market's insurance in force (within a firm-year, equivalent to the log number of lives, since average policy size is a firm-year constant) or its deciles; the Poisson column regresses losses on log distance with log premium as exposure, without a trim. The last row regresses log insurance in force on log distance. Firm-by-year effects; standard errors clustered by firm and state-year. Panel B repeats Figure IV, Panel C: coefficients on years-since-entry bins relative to founding markets, with distance bands controlled, firm-by-year effects, and firm clusters; the common sample has positive insurance in force; the Poisson columns use losses with log premium as exposure, in log points. Panel C re-estimates Table VIII with deciles of the market's log insurance in force added to the position term; its baseline columns retain one fewer firm-year at the trim boundary than Table VIII and differ from it by up to 0.2 points. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure IA.5: The loss ratio of entered markets with and without a control for market size



Coefficients and 95 percent confidence intervals from Table [IA.9a](#), Panel B: the loss ratio of entered markets by years since entry, relative to the firm's founding markets, as in Figure [IV](#), Panel C, and with the log of the market's insurance in force as a control.

Table IA.10: Within-market comparison: pooled estimates and weighting

Outcome	Breadth coefficient				Controls		$E[Y]$	N																														
	Base	Adding share	Adding tenure	Adding both	Within-state share	State tenure																																
<i>Panel A. Each market weighted equally</i>																																						
Loss ratio	0.029*** (0.009)	0.029*** (0.009)	0.027*** (0.007)	0.027*** (0.007)	0.034 (0.107)	0.0037*** (0.0006)	0.273	120,151																														
Claims per 1,000 in force	0.733*** (0.244)	0.749*** (0.247)	0.665*** (0.200)	0.698*** (0.203)	6.395** (2.915)	0.1225*** (0.0154)	8.243	117,852																														
Premium yield	-0.614*** (0.235)	-0.536** (0.238)	-0.633*** (0.231)	-0.548** (0.233)	15.945*** (4.022)	0.0261 (0.0224)	30.264	116,639																														
<i>Panel B. Weighted to reproduce the firm aggregate</i>																																						
Loss ratio	0.062*** (0.012)	0.055*** (0.012)	0.061*** (0.011)	0.054*** (0.012)	-0.343*** (0.077)	0.0010 (0.0008)	0.329	120,151																														
Claims per 1,000 in force	1.595*** (0.304)	1.473*** (0.306)	1.503*** (0.284)	1.376*** (0.284)	-6.638*** (2.079)	0.0326 (0.0211)	10.529	117,852																														
Premium yield	-0.527 (0.810)	-0.319 (0.872)	-0.749 (0.636)	-0.545 (0.692)	10.734** (4.742)	0.0125 (0.0313)	31.931	116,639																														
<table><tr><td></td><td>Equal, base</td><td>Equal, both</td><td>Weighted, base</td><td>Weighted, both</td></tr><tr><td colspan="5"><i>Panel C. Claims per \$1,000 in force by period (% of mean)</i></td></tr><tr><td>1888–1940</td><td>8.9*** (3.0)</td><td>8.5*** (2.5)</td><td>15.1*** (2.9)</td><td>13.1*** (2.7)</td></tr><tr><td>1888–1905</td><td>26.7*** (5.9)</td><td>24.1*** (6.1)</td><td>31.4*** (8.1)</td><td>31.8*** (8.0)</td></tr><tr><td>1906–1920</td><td>17.4*** (5.6)</td><td>11.5** (4.5)</td><td>28.7*** (6.3)</td><td>25.3*** (6.0)</td></tr><tr><td>1921–1940</td><td>2.7 (3.0)</td><td>3.6 (2.6)</td><td>11.5*** (3.0)</td><td>10.5*** (2.6)</td></tr></table>										Equal, base	Equal, both	Weighted, base	Weighted, both	<i>Panel C. Claims per \$1,000 in force by period (% of mean)</i>					1888–1940	8.9*** (3.0)	8.5*** (2.5)	15.1*** (2.9)	13.1*** (2.7)	1888–1905	26.7*** (5.9)	24.1*** (6.1)	31.4*** (8.1)	31.8*** (8.0)	1906–1920	17.4*** (5.6)	11.5** (4.5)	28.7*** (6.3)	25.3*** (6.0)	1921–1940	2.7 (3.0)	3.6 (2.6)	11.5*** (3.0)	10.5*** (2.6)
	Equal, base	Equal, both	Weighted, base	Weighted, both																																		
<i>Panel C. Claims per \$1,000 in force by period (% of mean)</i>																																						
1888–1940	8.9*** (3.0)	8.5*** (2.5)	15.1*** (2.9)	13.1*** (2.7)																																		
1888–1905	26.7*** (5.9)	24.1*** (6.1)	31.4*** (8.1)	31.8*** (8.0)																																		
1906–1920	17.4*** (5.6)	11.5** (4.5)	28.7*** (6.3)	25.3*** (6.0)																																		
1921–1940	2.7 (3.0)	3.6 (2.6)	11.5*** (3.0)	10.5*** (2.6)																																		

The unit of observation is the firm-state-year, ordinary insurance, 1888–1940. All columns estimate equation (5.3) with state $\times$ year fixed effects and a cubic in standardized log firm premium, with standard errors clustered by firm and by state-year. Breadth is measured at the firm-year level and is constant across an insurer's markets. Within-state share is the insurer's premium in a state divided by all insurers' premium in that state that year. $E[Y]$ reports the mean of each row's outcome on its own estimation sample. Panels A and B report coefficients in levels, pooled over 1888–1940; in Panel B markets are weighted to reproduce the firm aggregate. Panel C reports the claims coefficient by period in percent of the (weighted) mean, equally weighted or weighted by insurance in force, without controls (base) and with local share, tenure, and the founding-market indicator (both). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table IA.11: Within markets: further specifications
IA.11a. The level term: further specifications

	Loss ratio	Claims per \$1,000	Premium per \$1,000
<i>Panel A. Level term, each market weighted equally</i>			
Table IX, share + tenure, firm clusters	0.027*** (0.007)	0.698*** (0.205)	-0.548** (0.235)
two-way clustered, firm and state-year (as Table IX)	0.027*** (0.007)	0.698*** (0.203)	-0.548** (0.233)
+ firm-age bins, incumbent indicator	0.025*** (0.008)	0.610*** (0.212)	-0.635** (0.247)
+ HQ-state $\times$ year FE	0.017** (0.008)	0.431** (0.209)	-0.428 (0.262)
excluding the five largest insurers	0.026*** (0.008)	0.676*** (0.208)	-0.510** (0.238)
<i>Panel B. Level term, weighted to the firm aggregate</i>			
Table IX, share + tenure, firm clusters	0.054*** (0.012)	1.376*** (0.285)	-0.545 (0.700)
two-way clustered, firm and state-year (as Table IX)	0.054*** (0.012)	1.376*** (0.284)	-0.545 (0.692)
+ firm-age bins, incumbent indicator	0.054*** (0.012)	1.385*** (0.285)	-0.520 (0.704)
+ HQ-state $\times$ year FE	0.057*** (0.013)	1.653*** (0.337)	-0.113 (0.528)
excluding the five largest insurers	0.043*** (0.011)	1.176*** (0.284)	-0.155 (0.436)
<i>Panel C. Expansion elsewhere, incumbent markets</i>			
Table IA.12a, Panel C, col. 4, firm clusters	0.029*** (0.009)	0.811*** (0.272)	-0.158 (0.305)
+ tenure bins and firm-age bins	0.022*** (0.009)	0.585** (0.261)	-0.145 (0.315)
two-way clustered	0.022*** (0.008)	0.585** (0.259)	-0.145 (0.312)
years before 1907	0.001 (0.013)	-0.456 (0.498)	-0.991 (0.677)
<i>Panel D. Robertson Law: claims outside Texas of insurers that withdrew</i>			
Left Texas after 1907 $\times$ post	-0.011 (0.014)	-0.302 (0.483)	1.023 (1.281)

Panels A and B repeat the “adding both” column of Online Appendix Table IA.10. Standard errors are clustered by firm, except in the rows marked two-way clustered; Tables IX and IA.12a cluster by firm and by state-year, so their standard errors are those of the two-way rows (for Table IA.12a, column 4 uses linear tenure). Panel C repeats column 4 of Table IA.12a, Panel C. Panel D compares, in markets other than Texas held by 1906, insurers that wrote business in Texas in 1906 but no longer did so in 1909–1912 with those that continued, 1900–1914, with firm-state and state-by-year effects.

Table IA.12: Earlier firm-level specifications: firm fixed effects, New York controls, and lines of business

IA.12a. Firm fixed effects, New York controls, lines of business, and incumbent markets

<i>Panel A. Fixed effects and New York controls</i>				
	Region-by-year FE (1)	Firm FE (2)	NY sample (3)	NY + controls (4)
Premium yield	-1.007*** (0.256)	-2.827*** (0.614)	-0.977* (0.580)	-0.748 (0.536)
Loss ratio	0.036*** (0.008)	0.040*** (0.009)	0.031 (0.028)	0.039 (0.026)
Observations	7,629	7,586	1,864	1,762

<i>Panel B. By insurance type and product-mix controls</i>				
	Ordinary (1)	Industrial (2)	Group (3)	Pooled + mix (4)
Premium yield	-1.007*** (0.256)	-6.055*** (2.171)	-0.261 (0.710)	-1.065*** (0.341)
Loss ratio	0.036*** (0.008)	-0.004 (0.010)	-0.001 (0.030)	0.035*** (0.007)
Observations	7,629	1,253	794	7,832

<i>Panel C. The level term, within markets</i>				
	Baseline (1)	Adding log distance (2)	HQ region × year FE (3)	Expansion elsewhere (4)
Premium yield	-0.521** (0.240)	-0.408 (0.254)	-0.288 (0.256)	-0.158 (0.302)
Loss ratio	0.029*** (0.008)	0.037*** (0.008)	0.032*** (0.008)	0.029*** (0.009)
Claims per 1,000 in force	0.770*** (0.208)	0.981*** (0.215)	0.851*** (0.212)	0.811*** (0.269)
Observations	117,643	116,559	116,559	109,497

Notes. Each cell reports the coefficient on standardized breadth, the log effective number of states $N_{\text{eff}} = 1/\text{HHI}$, from a separate regression of the row outcome conditional on a cubic in standardized log firm premium, with the standard error below it. Standard errors are clustered by firm (in Panel C, by firm and by state-year) and single-state firm-years are excluded. Premium yield is premiums per \$1,000 of insurance in force, the loss ratio is claims over premiums, and claims per \$1,000 is claims over insurance in force. Zero-claim observations are retained and outcomes are trimmed at the top, with premium yield trimmed at both tails. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table IA.13: The insurers' own accounts: death claims, deaths, and actual and expected mortality

	1888–1905	1906–1920	1921–1926
<i>Panel A. Insurers' own death claims per \$1,000 in force (% of mean)</i>			
All insurers	15.4* (9.0)	17.5*** (4.9)	13.8*** (3.7)
Licensed in New York	18.8* (9.9)	1.7 (13.5)	8.7 (17.5)
Not licensed in New York	-4.1 (13.6)	17.3*** (4.7)	11.8*** (3.6)
Spectator state-table claims, same firm-years	23.3*** (7.3)	18.8*** (5.7)	14.7*** (4.4)
Own death claims, premium-weighted	12.8 (10.5)	7.0 (6.6)	6.1 (5.9)
State-table claims, premium-weighted, same firm-years	30.3*** (8.1)	23.8*** (5.6)	22.7*** (5.7)
Own death claims, controlling for endowment share and policy size	14.3 (11.6)	15.9*** (4.5)	13.3*** (3.7)
<i>Panel B. How often insured lives died, and how large the claims were</i>			
Deaths per 1,000 policies in force	17.0** (6.6)	15.8*** (4.4)	15.9*** (3.1)
Deaths per 1,000 policies, premium-weighted	16.8** (8.0)	10.0* (5.8)	16.6*** (5.3)
Face amount terminated by death per \$1,000	20.7*** (6.8)	16.1*** (4.6)	13.9*** (3.1)
Claim size relative to the average policy (log points)	4.6** (2.0)	0.1 (1.2)	-1.6 (1.1)
<i>Panel C. Actual and expected mortality (gain-and-loss exhibits)</i>			
A/E, all insurers reporting it	0.039 (0.057)	0.053*** (0.014)	0.046*** (0.010)
Insurers already reporting by 1910	—	0.019 (0.034)	0.024 (0.031)
Not licensed in New York	-0.092 (0.097)	0.062*** (0.015)	0.041*** (0.010)
Licensed in New York	0.010 (0.071)	-0.010 (0.041)	-0.001 (0.043)
Weighted by expected mortality	-0.116 (0.091)	-0.056* (0.033)	0.025 (0.018)
Actual mortality per \$1,000 in force (log points)	20.2** (9.5)	16.1*** (3.2)	14.1*** (2.7)
Expected mortality per \$1,000 in force (log points)	2.2 (5.5)	4.4*** (1.3)	3.9*** (1.3)
Claims level term on the A/E firm-years (% of mean)	39.7*** (9.6)	19.8*** (5.0)	—
controlling for A/E	38.7*** (10.3)	14.3*** (5.1)	—
<i>Panel D. Selection, agency, and policy mix (% of mean unless noted)</i>			
Medical fees per \$1,000 of new insurance	13.5 (8.3)	8.8*** (2.9)	6.4* (3.4)
Lapses per \$1,000 in force	17.1* (9.6)	14.8*** (4.9)	14.2*** (4.2)
Policies not taken, per \$100 of new issues	41.8*** (10.9)	—	—
Endowment share of insurance in force (points)	-3.5 (5.1)	-0.8 (1.3)	-1.1 (1.5)
Firm-years (Panel A, row 1)	598	2,243	899
Insurers	81	297	246
Firm-years (Panel C, row 1)	362	1,475	1,170
Insurers	60	227	246

Notes. Firm-years of the preferred sample of Table VIII, built separately for 1888–1920 and 1921–1940 (breadth and size standardized within each), matched to the company accounts and gain-and-loss exhibits of the *Insurance Year Book* (Online Appendix Section A.10). Each cell is the coefficient on breadth, conditional on a cubic in size, with headquarters-region-by-year effects and standard errors clustered by firm. Panels A, B, and D and rows 6 and 7 of Panel C use firm-years without industrial business whose company insurance in force is within a factor of 1.5 of the state-table sum; Panel C, rows 1–5, 8, and 9, use every firm-year with a gain-and-loss exhibit. Insurance in force is the mean of the current and previous year-ends. Panel A: death claims paid per \$1,000 in force, by whether the insurer filed a New York statement that year; row 4 is the Spectator's state-table losses per \$1,000 of state-table insurance in force on the firm-years of row 1; rows 5 and 6 weight rows 1 and 4 by premium and are in percent of the premium-weighted mean. Panel B: policies terminated by death per 1,000 policies in force (row 2 weighted by premium), the face amount terminated by death per \$1,000 in force, and the log of the average claim over the average policy in force. Panel C: A/E is actual over expected mortality; row 2 keeps the insurers whose first exhibit is for 1910 or earlier; row 5 weights firm-years by expected mortality; rows 6 and 7 are log mortality per \$1,000 in force (times 100); rows 8 and 9 are the claims level term of Table VIII in percent of the mean claims aggregate, without and with A/E as a control. The mean A/E is 0.76, 0.64, and 0.49 in the three periods, and mean death claims per \$1,000 are 11.3, 6.8, and 5.4. Panel D: medical fees and inspections per \$1,000 of new insurance (from 1905), lapses per \$1,000 in force, policies not taken per \$100 of insurance issued (reported to 1907), and the endowment share of insurance in force (times 100). Outcomes are trimmed at the 1st and 99th percentiles within each sample. Coefficients are in percent of the outcome mean unless noted; A/E rows in levels. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table IA.14: The level term in the New York statements

	1888–1905	1906–1920
<i>Panel A. Claims per \$1,000 in force on breadth (% of mean)</i>		
Spectator losses, firm aggregate	26.3*** (8.7)	17.0 (11.8)
Spectator losses, level term	36.3*** (8.1)	23.4* (13.5)
NY statements, death claims	21.0*** (8.0)	4.4 (13.6)
<i>Panel B. Seasoning and policy mix</i>		
Reserve per \$1,000 on firm age (per year)	13.2*** (1.9)	6.1*** (1.7)
Reserve per \$1,000 on breadth, unconditional (% of mean)	12.0 (10.3)	14.5** (5.8)
Reserve per \$1,000 on breadth (% of mean)	-12.7 (11.0)	5.2 (9.4)
Endowment share of insurance in force on breadth (points)	-18.5** (9.3)	-0.2 (8.5)
NY death claims, controlling for reserve and endowment share	19.2** (9.0)	0.2 (9.8)
NY death claims, controlling for firm age	20.5** (8.1)	2.5 (14.9)
<i>Panel C. Participation</i>		
Policyholder dividends per \$100 of premium on breadth	-1.9 (1.7)	1.5 (1.9)
Spectator level term, controlling for the dividend rate	40.3*** (7.7)	20.1 (13.9)
<i>Panel D. Tabular mortality, insurers reporting A/E (log points)</i>		
Actual mortality per \$1,000 in force	—	-5.1 (12.2)
Expected mortality per \$1,000 in force	—	-7.8 (7.6)
Actual / expected mortality	—	2.6 (6.4)
Firm-years (Panel A, row 3)	454	458
Insurers	41	41

Notes. Firm-years of the 1888–1920 sample of Table VIII matched to the ordinary department of the insurers' New York statements, where the statement's ordinary insurance in force is within a factor of 1.5 of the Spectator's state sum. Each cell is the coefficient on breadth, conditional on a cubic in size, with headquarters-region-by-year effects and standard errors clustered by firm, except the first row of Panel B, which regresses the reserve on firm age and log premium with year effects. Panel A: the Spectator's ordinary losses per \$1,000 of insurance in force (the firm aggregate and the level term of the exact decomposition) and death claims per \$1,000 in force from the policy exhibit of the statements, both in percent of the mean. Panel B: the policy reserve ("reinsurance reserve") per \$1,000 of ordinary insurance in force, which rises with the duration of the policies in force and with the endowment share; the endowment share of insurance in force in percentage points; and the death-claims coefficient of Panel A with the stated controls. Panel C: policyholder dividends per \$100 of premium income, and the Spectator level term controlling for it. Panel D: insurers without industrial business that report A/E (Online Appendix Table IA.20a); actual and expected mortality are on the net amount at risk, per \$1,000 of ordinary insurance in force, in logs (times 100), so the first row equals the sum of the other two. Death claims, reserves, and the dividend rate are trimmed at the 1st and 99th percentiles. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table IA.15: Own death claims and deaths: robustness
IA.15a. The insurers' own accounts: robustness

	1888–1905	1906–1920	1921–1926
<i>Panel A. Own death claims per \$1,000 (% of mean)</i>			
Baseline	15.4* (9.0)	17.5*** (4.9)	13.8*** (3.7)
Year instead of region-by-year effects	15.1 (9.9)	20.1*** (5.8)	16.9*** (4.6)
Weighted (premium; for A/E, expected mortality)	12.8 (10.5)	7.0 (6.6)	6.1 (5.9)
Without firm-years with a jump of more than threefold	13.9 (9.5)	17.8*** (5.0)	14.5*** (3.9)
Controlling for firm age	13.1 (9.9)	15.7*** (4.0)	13.8*** (3.2)
Controlling for new business / insurance in force	16.0* (9.0)	18.8*** (4.5)	16.1*** (3.4)
Logarithm of the outcome (log points)	23.3*** (7.4)	16.4*** (4.2)	12.4*** (3.2)
<i>Panel B. Deaths per 1,000 policies (% of mean)</i>			
Baseline	15.8** (7.4)	16.0*** (4.4)	15.7*** (3.1)
Year instead of region-by-year effects	20.3*** (6.7)	17.7*** (5.1)	17.4*** (3.7)
Weighted (premium; for A/E, expected mortality)	16.3** (8.2)	10.2* (5.8)	16.6*** (5.3)
Without firm-years with a jump of more than threefold	13.7* (7.5)	17.0*** (4.6)	15.7*** (3.4)
Controlling for firm age	14.2* (8.1)	14.5*** (3.8)	16.0*** (2.9)
Controlling for new business / insurance in force	17.1** (7.0)	16.7*** (4.3)	17.7*** (2.9)
Logarithm of the outcome (log points)	18.6** (7.3)	16.4*** (3.8)	14.5*** (2.8)
<i>Panel C. A/E (ratio)</i>			
Baseline	0.039 (0.057)	0.053*** (0.014)	0.046*** (0.010)
Year instead of region-by-year effects	0.058 (0.051)	0.053*** (0.016)	0.047*** (0.011)
Weighted (premium; for A/E, expected mortality)	-0.116 (0.091)	-0.056* (0.033)	0.025 (0.018)
Without firm-years with a jump of more than threefold	0.032 (0.058)	0.056*** (0.014)	0.045*** (0.010)
Controlling for firm age	0.039 (0.058)	0.050*** (0.016)	0.045*** (0.010)
Controlling for new business / insurance in force	0.051 (0.049)	0.051*** (0.014)	0.048*** (0.010)
Logarithm of the outcome (log points)	8.1 (7.2)	9.9*** (2.3)	9.3*** (1.9)
<i>Panel D. A/E on the firm-years of Table IA.20a</i>			
New York A/E, all firm-years of Table IA.20a	—	-0.063 (0.045)	—
New York A/E, firm-years with a Spectator exhibit	—	-0.011 (0.041)	—
Spectator A/E, the same firm-years	—	-0.008 (0.041)	—

Panels A–C re-estimate Table IA.13, Panel A, row 1, Panel B, row 1, and Panel C, row 1. Weights are premium (Panels A and B) and expected mortality (Panel C); weighted rows are in percent of the weighted mean of the outcome. The jump screen drops firm-years whose company insurance in force, number of policies in force, death claims, or premium differs from the previous year by more than a factor of three. New business is the amount of new policies issued over the mean insurance in force. Panel D: 1906–1920 firm-years of Table IA.20a, Panel A, column 3; row 1 reproduces that estimate, row 2 restricts it to the firm-years with a Spectator gain-and-loss exhibit, and row 3 uses the Spectator's A/E on those firm-years. Size cubic, region-by-year effects (except row 2 of each panel), firm clusters. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table IA.16: Expansion and claims
IA.16a. Timing of entry and prior claims experience

	(1)	(2)	(3)	(4)
Loss ratio, year t	-1.96*** (0.64)	-1.88*** (0.66)	0.41 (0.68)	0.89 (0.84)
one year earlier				-0.44 (0.73)
two years earlier				-0.97* (0.54)
three years earlier				0.46 (0.71)
New business / in force		4.09*** (0.73)	2.73*** (0.98)	5.44*** (1.22)
Log effective states	1.99** (0.97)	1.53 (1.01)	-10.69*** (1.82)	-10.40*** (1.88)
Share of home-state market		1.05 (0.85)		
Log firm premium	1.37 (1.29)	2.38 (1.49)	1.77 (3.26)	2.24 (3.86)
Firm age	-6.15*** (0.89)	-4.96*** (0.87)		
Firm fixed effects	No	No	Yes	Yes
Sum of loss-ratio terms				-0.06
p-value, loss-ratio terms jointly				0.396
Mean of dependent variable	14.91	15.16	15.00	13.57
Observations	7,357	6,847	6,933	5,400
Adjusted R-squared	0.055	0.070	0.146	0.145

Notes. This table estimates an insurer's entry into a new state on its characteristics in preceding years. The dependent variable equals 100 if the firm opens at least one new state in year $t + 1$ and zero otherwise, where an entry is a firm-state observation in its first year with positive premium, new insurance written exceeding half of insurance in force, and entry premium no greater than the firm's median state premium that year. The unit of observation is the firm-year. Every regressor is standardized to zero mean and unit variance over the firm-year panel, so a coefficient is the change in the probability of entry next year, in percentage points, per standard deviation of the regressor. Standard errors, in parentheses, are clustered by firm. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table IA.16: Expansion and claims (continued)
IA.16b. Incumbent markets around an expansion burst

Event time	Claims per \$1,000			Loss ratio		
	Two-way FE	Pre-trend removed	Matched, stacked	Two-way FE	Pre-trend removed	Matched, stacked
$t - 5$	-1.18 (0.72)	0.35 (0.28)	-3.44*** (0.78)	-0.034 (0.022)	0.014 (0.012)	-0.100*** (0.024)
$t - 4$	-0.75 (0.50)	0.49* (0.28)	-2.29*** (0.86)	-0.028* (0.016)	0.016 (0.010)	-0.073*** (0.022)
$t - 3$	-1.07** (0.47)	0.22 (0.38)	-1.68*** (0.55)	-0.038** (0.017)	0.006 (0.008)	-0.050** (0.021)
$t - 2$	-0.35 (0.32)	0.74 (0.49)	-1.15** (0.52)	-0.017 (0.013)	0.025* (0.014)	-0.056*** (0.021)
$t - 1$	ref.	1.67*** (0.58)	ref.	ref.	0.062*** (0.017)	ref.
t	0.35 (0.36)	1.43** (0.56)	-0.40 (0.58)	-0.002 (0.013)	0.028 (0.017)	-0.022 (0.020)
$t + 1$	0.33 (0.39)	1.16** (0.56)	-1.02 (0.68)	0.015 (0.014)	0.037** (0.018)	-0.003 (0.020)
$t + 2$	0.09 (0.44)	1.37* (0.80)	-1.37* (0.73)	0.002 (0.014)	0.040* (0.023)	-0.026 (0.024)
$t + 3$	0.46 (0.45)	1.21* (0.67)	-2.08** (0.83)	0.006 (0.015)	0.031 (0.020)	-0.092** (0.038)
$t + 4$	0.75 (0.47)	1.60** (0.78)	-2.12** (0.96)	0.020 (0.014)	0.043 (0.030)	-0.119** (0.048)
$t + 5$	0.08 (0.51)	0.78 (0.89)	-2.39*** (0.84)	0.003 (0.018)	0.015 (0.028)	-0.083*** (0.032)
$t + 6$	0.89* (0.53)	1.66 (1.19)	-0.51 (1.00)	0.030 (0.019)	0.015 (0.040)	0.018 (0.029)
Joint test of leads, p -value	0.219	0.011	0.000	0.283	0.003	0.000
Observations	54,036	3,731	15,968	54,867	3,914	17,344

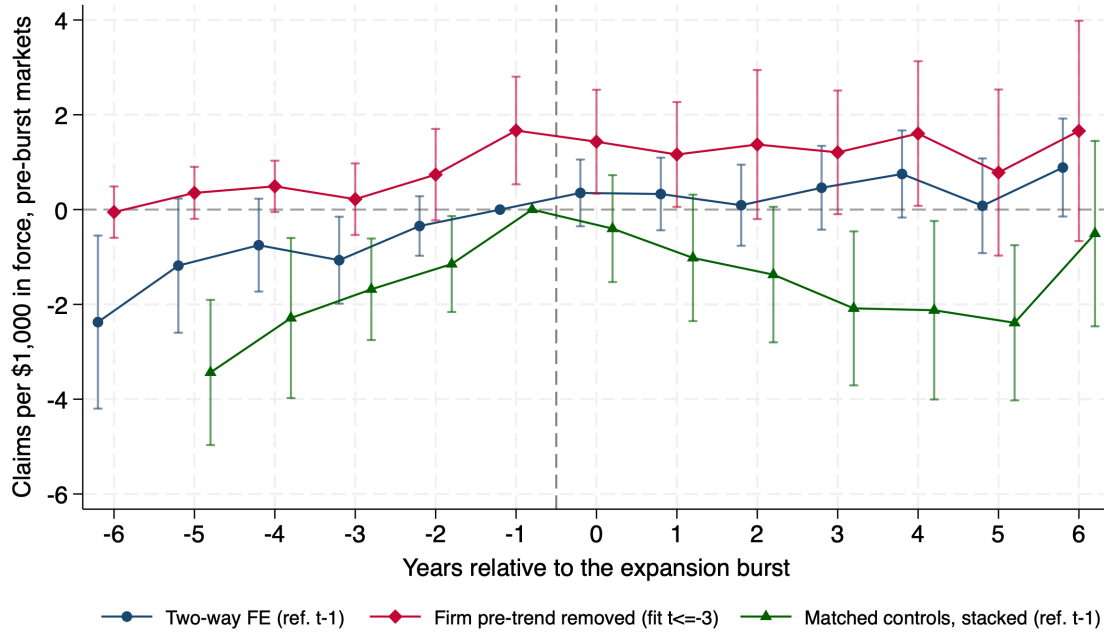
A burst is the first year after an insurer's first observed year in which it enters three or more new states; outcomes are measured in the markets it held before the burst, excluding its first year in each. *Two-way FE*: event-time indicators (endpoints binned, $t - 1$ omitted), firm-state and state-by-year effects, tenure bins, local share, and a cubic in size; never-bursting insurers are controls. *Pre-trend removed*: the same effects and controls are estimated on never-bursting insurers and on bursting insurers' years up to $t - 3$; each bursting insurer's residual linear trend over those years is removed, and $t - 2$ and $t - 1$ are out-of-sample placebos. *Matched, stacked*: each bursting insurer with up to three never-bursting insurers closest in size and firm age at $t - 1$, over $t - 5$ to $t + 6$, with stack-specific firm-state and state-by-year effects. Standard errors clustered by firm; the joint test covers the leads $t - 5$ to $t - 2$ (for the trend design, $t - 2$ and $t - 1$).

Table IA.17: Claims in incumbent markets around expansion episodes: stacked design

	Baseline	New-business control	Episodes 1906 or earlier	Episodes after 1906
<i>Panel A. Claims per \$1,000 in incumbent markets, by year relative to the episode</i>				
-5	-0.609 (0.553)	-0.422 (0.529)	-2.155 (2.156)	-0.438 (0.567)
-4	-0.509 (0.430)	-0.353 (0.440)	-1.805 (1.152)	-0.363 (0.452)
-3	-1.031*** (0.380)	-0.805** (0.395)	-3.724*** (1.137)	-0.743** (0.343)
-2	-0.376 (0.320)	-0.141 (0.339)	-0.651 (1.306)	-0.358 (0.321)
0	0.096 (0.369)	0.234 (0.372)	-1.125 (1.360)	0.102 (0.376)
1	-0.026 (0.319)	0.249 (0.323)	-1.853 (1.219)	0.078 (0.334)
2	-0.351 (0.459)	-0.116 (0.414)	-3.193 (1.944)	-0.077 (0.441)
3	-0.256 (0.423)	-0.004 (0.398)	-3.128** (1.452)	0.077 (0.404)
4	0.297 (0.530)	0.540 (0.487)	-3.446* (2.049)	0.840* (0.467)
5	-0.701 (0.561)	-0.439 (0.540)	-4.627** (1.879)	-0.144 (0.553)
6	-0.403 (0.617)	-0.204 (0.578)	-6.004*** (1.732)	0.657 (0.572)
<i>Panel B. Pooled coefficients (claims per \$1,000)</i>				
Years 0 to 6	-0.170 (0.367)	0.058 (0.340)	-3.197** (1.508)	0.186 (0.346)
Years -5 to -2	-0.626** (0.301)	-0.418 (0.311)	-1.953 (1.189)	-0.479* (0.291)
Joint test of years -5 to -2 (<i>p</i> -value)	0.116	0.331	0.012	0.292
Observations	137,186	135,072	21,144	116,042
<i>Panel C. Pooled years 0 to 6 by number of states entered in the episode year</i>				
Three or four states	-0.184 (0.429)			
Five or more states	-0.143 (0.568)			
Treated insurer-episodes	97			
Mean before the episode	6.60			

Stacked event study, equation (5.4). An expansion episode is the first year after an insurer's first appearance in which it enters three or more states. For each episode year c , the treated insurers are those whose first episode occurs in c , and the controls are insurers with no episode or with a first episode after $c + 6$; all are active in $c - 1$ in at least two states. The sample consists of their incumbent markets (states served before c , excluding each market's first year) over $c - 5$ to $c + 6$. The outcome is claims per \$1,000 of insurance in force, trimmed at the 97th percentile. All specifications include market-by-stack and state-by-year-by-stack fixed effects, a cubic in size, the insurer's share of the state's premium, and tenure indicators; year -1 is omitted. Column 2 adds the insurer's ratio of new insurance written to insurance in force; columns 3 and 4 restrict to episodes in 1906 or earlier and after 1906. Panel B pools the years 0 to 6 and -5 to -2 . Panel C pools the years after the episode separately for episodes entering three or four states and five or more. Standard errors, clustered by firm, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Figure VIII plots columns 1 and 2 of Panel A.

Figure IA.6: Claims in incumbent markets around an expansion burst



Notes. Claims per \$1,000 in force in the markets an insurer held before its first burst of expansion, by year relative to the burst, under the three designs of Online Appendix Table [IA.16b](#). 95 percent confidence intervals, clustered by firm.

Table IA.18: Business model: deferred dividends, participation, organization, income, and expenses
IA.18a. Deferred-dividend writers

	Claims per Aggregate	\$1,000 Level	Loss ratio	Premium per \$1,000	N
<i>Panel A. Preferred specification, with and without the deferred-dividend writers</i>					
1888–1906: all insurers	18.6*** (6.7)	25.1*** (7.0)	25.7*** (7.1)	-15.0*** (3.8)	1,132
without deferred-dividend writers	16.6** (7.1)	23.6*** (7.4)	24.8*** (7.8)	-16.1*** (4.1)	996
without the Equitable, Mutual and New York Life	18.9*** (6.9)	25.6*** (7.2)	26.3*** (7.4)	-15.5*** (4.0)	1,081
breadth × deferred-dividend writer	16.7 (18.1)		15.4 (17.5)	3.2 (9.8)	1,132
1907–1920: all insurers	16.7*** (4.6)	18.8*** (4.2)	19.8*** (4.6)	-2.8** (1.2)	3,031
without deferred-dividend writers	17.0*** (4.9)	19.4*** (4.5)	20.1*** (5.0)	-2.8** (1.2)	2,919
without the Equitable, Mutual and New York Life	17.8*** (4.8)	20.1*** (4.4)	20.3*** (4.8)	-2.6** (1.2)	2,989
1888–1920: all insurers	18.7*** (4.4)	21.8*** (4.3)	22.6*** (4.5)	-5.6*** (1.3)	4,162
without deferred-dividend writers	18.3*** (4.7)	22.0*** (4.5)	22.5*** (5.0)	-5.9*** (1.3)	3,914
<i>Panel B. New York statements: controlling for the deferred-dividend liability</i>					
1906–1920					
Base (common sample)	9.4 (13.7)		6.0 (13.9)	1.4 (2.2)	518
+ deferred-dividend share of liabilities	7.6 (13.7)		6.0 (14.2)	-0.1 (2.0)	518
+ and policyholder-dividend payout	4.2 (13.5)		3.4 (14.3)	-0.6 (1.8)	518
1906–1940					
Base (common sample)	-1.3 (7.4)		-3.7 (8.3)	0.2 (1.6)	1,340
+ deferred-dividend share of liabilities	-1.2 (7.4)		-3.2 (8.4)	-0.1 (1.6)	1,340
+ and policyholder-dividend payout	-2.7 (6.7)		-4.5 (7.9)	-0.4 (1.5)	1,340

Notes. Deferred-dividend writers are the Equitable, the Mutual, and New York Life, and the insurers whose deferred-dividend liability was at least 5 percent of total liabilities in the first year the New York statements report it (Northwestern Mutual, Penn Mutual, Union Central, Germania, and National of Vermont). Panel A: the specification of Table VIII with breadth and size standardized within each period; the interaction row adds breadth times an indicator for these firms and the indicator itself. Panel B: New York statement firm-years of the sample of Table VIII, Panel C, with the deferred-dividend liability over total liabilities and policyholder dividends over premium as controls, on a common sample. Coefficients in percent of the outcome mean; region-by-year effects; firm clusters. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table IA.18: Business model: deferred dividends, participation, organization, income, and expenses (continued)

IA.18b. Organizational form, dividends, expenses, and the Armstrong legislation

	Breadth	Breadth + mutual	Mutual	Breadth × mutual	$E[Y]$
<i>Panel A. New York accounting sample</i>					
Loss ratio (Spectator)	0.031 (0.028)	0.030 (0.028)	0.021 (0.032)	0.005 (0.033)	0.377
Premium per \$1,000	-0.977* (0.580)	-0.990* (0.586)	0.186 (0.829)	-1.061 (1.029)	32.470
Claims per \$1,000	0.735 (0.823)	0.688 (0.825)	0.639 (0.988)	-0.281 (1.006)	12.100
Losses / premiums (NY)	0.010 (0.028)	0.010 (0.029)	-0.002 (0.033)	-0.004 (0.034)	0.400
(Losses + policyholder dividends) / premiums	0.027 (0.033)	0.024 (0.033)	0.033 (0.036)	0.017 (0.038)	0.505
Policyholder dividends / premiums	0.015 (0.009)	0.012 (0.008)	0.041*** (0.013)	0.016 (0.014)	0.105
(Commissions + salaries) / premiums	-0.010 (0.012)	-0.008 (0.012)	-0.024 (0.015)	0.038 (0.034)	0.203
Commissions / premiums	0.005 (0.008)	0.006 (0.008)	-0.010 (0.008)	0.013 (0.019)	0.113
Lapse and surrender payments / premiums	-0.009 (0.013)	-0.010 (0.013)	0.006 (0.016)	-0.001 (0.017)	0.143
Deferred-dividend liability / total liabilities	0.005** (0.002)	0.005** (0.002)	0.001 (0.004)	0.001 (0.005)	0.016
<i>Panel B. Before and after the Armstrong legislation, full sample</i>					
	1888–1906	1907–1940			
Loss ratio (Spectator)	0.092*** (0.022)	0.029*** (0.007)			
Premium per \$1,000	-3.546*** (1.108)	-0.882*** (0.259)			
Claims per \$1,000	2.749*** (0.713)	0.663*** (0.207)			

Firm-years of multi-state insurers, with a cubic in size and region-by-year effects; standard errors clustered by firm. Mutual equals one when the New York statement reports no capital stock; the fourth column is the interaction of breadth and mutual status. Panel A ratios are from New York income statements, excluding industrial departments.

Table IA.19: Timing: New York-licensed and other insurers, and the years after 1920
IA.19a. The claims penalty of New York-licensed and other insurers, and where within the insurer's
markets it fell

	New York-licensed	Other insurers	All			
<i>Panel A. Claims level term on breadth by period (% of mean), 1888–1940 sample</i>						
1888–1895	26.7** (10.8)	2.5 (16.2)	19.2** (8.1)			
1896–1900	34.1*** (10.3)	-29.0 (20.1)	7.2 (10.3)			
1901–1905	41.0*** (14.4)	12.7 (13.3)	24.5** (11.6)			
1906–1910	30.6* (16.7)	30.3** (13.1)	31.2*** (8.8)			
1911–1915	6.1 (15.0)	20.2*** (7.6)	23.8*** (6.5)			
1916–1920	20.5 (15.1)	9.0** (4.0)	12.5*** (4.2)			
1921–1925	8.7 (15.0)	5.0 (5.8)	9.3* (5.2)			
1926–1930	-7.1 (7.9)	1.4 (4.7)	-0.9 (4.3)			
1931–1935	-14.9* (8.2)	11.0** (5.5)	5.6 (5.1)			
1936–1940	-5.4 (9.3)	2.7 (6.6)	1.0 (5.8)			
<i>Panel B. 1888–1905 and 1906–1920 (% of mean)</i>						
1888–1905	35.0*** (8.5)	2.5 (10.7)				
1906–1920	14.7 (14.9)	19.0*** (5.3)				
Breadth $\times$ post-1906 $\times$ New York-licensed (per \$1,000)			-2.33 (1.49)			
	1888–1905	LOO	1906–1920	LOO	1921–1940	LOO
<i>Panel C. Incidence within the insurer's markets: state-year effects (Table X) and a leave-own-firm-out benchmark</i>						
Breadth (nearest third)	3.65*** (0.81)	3.01*** (0.84)	1.30*** (0.44)	0.85* (0.45)	0.44** (0.22)	0.14 (0.23)
$\times$ headquarters state	-1.33** (0.67)	-1.38** (0.68)	-0.41 (0.27)	-0.19 (0.27)	-0.55*** (0.16)	-0.52*** (0.16)
$\times$ largest other market	-1.64 (1.05)	-1.60 (1.08)	-1.62*** (0.51)	-0.89* (0.47)	-1.07*** (0.27)	-0.68** (0.27)
$\times$ middle third	-1.15** (0.56)	-0.58 (0.60)	-0.09 (0.24)	0.25 (0.29)	0.10 (0.16)	0.23 (0.17)
$\times$ farthest third	-1.55** (0.77)	-0.78 (0.84)	0.07 (0.34)	0.59 (0.36)	0.09 (0.21)	0.25 (0.22)

Panels A and B: the specification of Figure VII (claims level term on breadth, size cubic, region-by-year effects, firm clusters, in percent of the period's mean claims) on firm-years in which the insurer filed a New York statement and on the rest; periods with fewer than ten insurers are omitted. The triple difference is estimated on 1888–1920 with breadth, breadth $\times$ post-1906, breadth $\times$ New York licensing, New York licensing, and post-1906 $\times$ New York licensing, in dollars per \$1,000. Panel C: the specification of Table X, Panel A, with state-year effects, and (LOO) with the outcome measured as the cell's claims per \$1,000 less those of the other insurers in the state and year, with year effects; per \$1,000. Table IX reports the within-market claims coefficient by period. Standard errors in Panel C are clustered by firm and by state-year. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table IA.19: Timing: New York-licensed and other insurers, and the years after 1920 (continued)
IA.19b. The claims gap after 1920

	1888–1905	1906–1920	1921–1926	
<i>Panel A. Claims per \$1,000 on breadth (% of mean), by sample</i>				
Table VIII sample	—	18.6*** (5.1)	2.1 (5.2)	
without firm-years above the 99th percentile of claims	—	19.1*** (4.9)	8.7** (3.9)	
level term, same firm-years	—	22.2*** (4.5)	8.2* (4.4)	
Firm-years with company accounts	—	17.4*** (5.0)	4.3 (5.3)	
Without insurers writing industrial business	—	19.1*** (5.7)	5.0 (6.3)	
Company in force within 1.5 times the state sum	—	20.3*** (5.5)	5.1 (5.0)	
Both restrictions (Table IA.13 sample)	—	21.7*** (6.0)	5.5 (5.9)	
without firm-years above the 99th percentile	—	20.8*** (5.7)	15.5*** (4.5)	
Own death claims, Table IA.13 sample	—	17.5*** (4.9)	13.8*** (3.7)	
	1921–1925	1926–1930	1931–1935	1936–1940
<i>Panel B. After 1920 by five-year period, 1921–1940 sample without the top percentile of claims</i>				
Claims aggregate	10.2** (4.0)	6.2 (4.0)	7.8** (3.8)	4.0 (4.2)
Claims level term	10.2** (4.8)	1.5 (4.0)	7.1 (5.0)	3.2 (5.5)

Panel A: claims per \$1,000 on breadth, in percent of the mean, on the preferred samples of Table VIII (1906–1920 from the 1888–1920 sample, 1921–1926 from the 1921–1940 sample), and on successive restrictions toward the sample of Table IA.13: firm-years with company accounts, without insurers writing industrial business in the year or an adjacent year, with company insurance in force within a factor of 1.5 of the state-table sum, and both. “Without firm-years above the 99th percentile” drops firm-years outside the 1st and 99th percentiles of claims within each sample. The level-term row is in percent of the mean of the claims aggregate. The last row is own death claims, Table IA.13, Panel A, row 1. Panel B: the 1921–1940 sample without the top and bottom percentiles of claims, by five-year period. Size cubic, region-by-year effects, firm clusters. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table IA.20: Actual-to-expected mortality
IA.20a. Actual versus expected mortality in the New York reports

<i>Panel A. Actual / expected mortality on breadth</i>							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Breadth (s.d.)	-0.046 (0.034)	-0.085* (0.044)	-0.063 (0.045)	-0.124** (0.051)	-0.032 (0.045)	0.008 (0.063)	-0.064 (0.041)
Years	1906–20	1906–20	1906–20	1906–20	1906–20	1906–08	1909–20
Size cubic	No	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects	Year	Year	Region-year	Region-year	Region-year	Region-year	Region-year
Sample, weights				Weighted	No industrial		
Mean A/E	0.732	0.732	0.734	0.734	0.719	0.728	0.735
Firm-years	521	521	517	517	472	108	409
Insurers	48	48	46	46	43	40	39

<i>Panel B. The claims cost on the same firm-years, 1906–1920, with A/E as a control</i>							
	Claims per \$1,000				Loss ratio		
	Level		Aggregate		Level		
Breadth (% of mean)	16.0 (15.1)	21.1 (13.1)	9.7 (13.9)	14.5 (12.0)	11.1 (15.3)	16.0 (13.4)	
A/E		11.23*** (4.35)		10.54** (4.13)		0.326** (0.134)	
Firm-years	517	517	517	517	517	517	

<i>Panel C. As Panel B, 1906–1908</i>							
	Claims per \$1,000				Loss ratio		
	Level		Aggregate		Level		
Breadth (% of mean)	45.3*** (16.6)	44.9*** (16.7)	31.1* (16.1)	30.8* (16.2)	41.2** (16.6)	40.9** (16.7)	
A/E		6.94 (5.91)		5.39 (5.59)		0.155 (0.177)	
Firm-years	108	108	108	108	108	108	

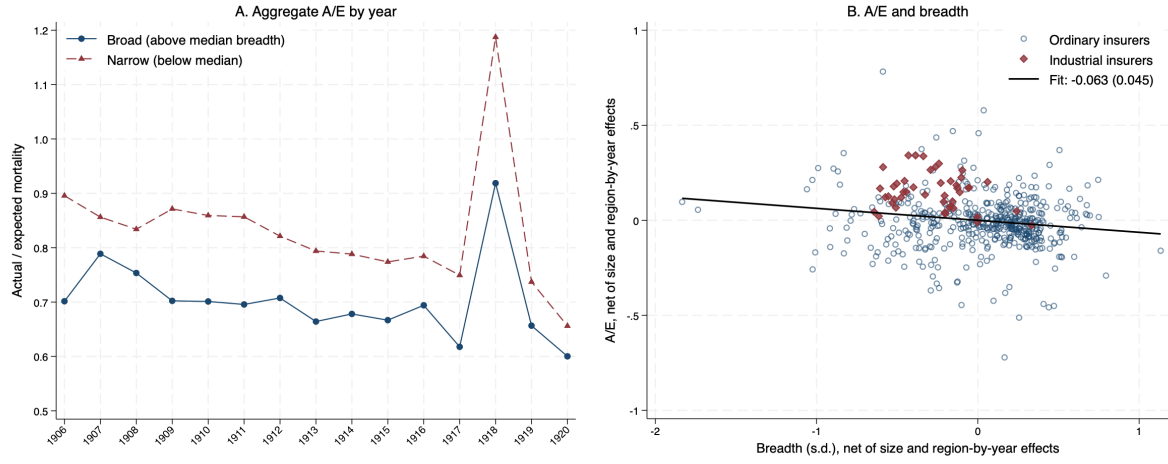
Notes. A/E is a company's actual mortality divided by the expected (tabular) mortality of its gain-and-loss exhibit, from the company statements in the New York Superintendent's reports for 1906–1908 and from their Table XI for 1909–1920 (Section A.9); the figures are company-wide and include industrial business. The sample is the firm-years of the 1888–1920 preferred sample of Table VIII with an A/E observation, excluding Bankers Life of Iowa before 1915, whose actual mortality then included amounts applied in reduction of assessments. Breadth and size are standardized over 1888–1920. Panel A reports regressions of A/E on breadth; column 4 weights firm-years by expected mortality; column 5 excludes the Metropolitan, Prudential, and John Hancock, whose business was mostly industrial; columns 6 and 7 split the period. Firm-years that are the only observation in their region-year are dropped from the region-by-year specifications. Panel B re-estimates Table VIII on the firm-years of column 3, with and without A/E as a control, and Panel C does so for 1906–1908; the breadth coefficient is expressed in percent of the mean of the outcome's aggregate, and the A/E coefficient in the outcome's units per unit of A/E. Standard errors are clustered by firm. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table IA.20: Actual-to-expected mortality (continued)
 IA.20b. Actual over expected mortality: reporting cohort, fixed effects, weights, and valuation basis

	1888–1905	1906–1920	1921–1926
<i>Panel A. Reporting cohort, fixed effects, and weights</i>			
Baseline (Table IA.13, Panel C, row 1)	0.039 (0.057)	0.053*** (0.014)	0.046*** (0.010)
Insurers already reporting by 1910	—	0.019 (0.034)	0.024 (0.031)
Years 1906–1911 only	—	0.028 (0.039)	—
Years 1912–1920 only	—	0.056*** (0.013)	—
Headquarters-state-by-year effects	-0.047 (0.074)	0.025 (0.019)	0.034*** (0.013)
Firm and year effects	-0.083 (0.054)	0.034 (0.022)	-0.029* (0.016)
Weighted by premium	-0.035 (0.080)	-0.030 (0.029)	0.024* (0.014)
Weighted by expected mortality	-0.116 (0.091)	-0.056* (0.033)	0.025 (0.018)
Below-median size	0.070 (0.067)	0.068*** (0.018)	0.047*** (0.017)
Above-median size	-0.033 (0.101)	0.025 (0.022)	0.038*** (0.012)
<i>Panel B. Valuation basis (1924 Reports of Life Insurance Companies)</i>			
Insurers with a known reserve basis	0.013 (0.065)	0.061*** (0.021)	0.047*** (0.013)
controlling for the reserve basis	0.002 (0.059)	0.056*** (0.019)	0.042*** (0.012)
and for mutual organization	-0.002 (0.062)	0.054*** (0.020)	0.041*** (0.012)
basis controls and headquarters-state-by-year effects	-0.048 (0.087)	0.042* (0.022)	0.033** (0.017)
basis controls, insurers reporting by 1910	—	0.011 (0.033)	0.053* (0.027)
Actuaries' table in the reserve basis, on breadth	0.212 (0.200)	0.195** (0.079)	0.107* (0.058)

Firm-years of Table IA.13, Panel C, row 1. *Reporting by 1910*: insurers whose first gain-and-loss exhibit in the sample is for 1910 or earlier. Size groups split at the median of size within the period among firm-years with an exhibit. Panel B uses the insurers linked to the 1924 *Reports of Life Insurance Companies* with a complete reserve basis (Online Appendix Section A.10); the basis controls are indicators for any business valued on the Actuaries' table, for preliminary-term valuation, and for full level-premium valuation, and the highest valuation interest rate; the mutual indicator is the form of organization reported there. The last row regresses the Actuaries'-table indicator on breadth. Size cubic, region-by-year effects unless noted, firm clusters. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure IA.7: The ratio of actual to expected mortality and breadth



Firm-years of Table [IA.20a](#). Panel A: aggregate A/E, the sum of actual over the sum of expected mortality, of the insurers above and below the year's median breadth. Panel B: A/E against breadth, both residualized on a cubic in size and region-by-year effects, with the fitted line (slope and firm-clustered standard error in the legend); diamonds are the three industrial insurers. The legend reports the estimate of Table [IA.20a](#), Panel A, column 3.

Table IA.21: Inference with few clusters and many tests
IA.21a. Inference with few clusters and many tests

	Estimate	t	p	Clusters / q
<i>Panel A. Wild-cluster bootstrap, New York samples</i>				
Own death claims, New York reports, 1888–1905 (% of mean)	21.6	2.82	0.035	41
Spectator claims, same firm-years, 1888–1905	26.9	3.26	0.013	41
Own death claims, New York reports, 1906–1920	4.4	0.33	0.754	41
Spectator claims, same firm-years, 1906–1920	17.0	1.49	0.149	41
A/E, New York reports, 1906–1920 (Table IA.20a, col. 3)	-0.063	-1.44	0.189	46
weighted by expected mortality (col. 4)	-0.124	-2.51	0.029	46
<i>Panel B. Company-account family (Table IA.13): conventional p and Benjamini–Hochberg q</i>				
Own death claims, 1888–1905	15.4	1.71	0.0875	0.0985
Own death claims, 1906–1920	17.5	3.59	0.0003	0.0005
Own death claims, 1921–1926	13.8	3.76	0.0002	0.0005
Deaths per 1,000 policies, 1888–1905	17.0	2.57	0.0103	0.0132
Deaths per 1,000 policies, 1906–1920	15.8	3.57	0.0004	0.0005
Deaths per 1,000 policies, 1921–1926	15.9	5.10	<0.0001	<0.0001
A/E, 1888–1905	0.039	0.69	0.4887	0.4887
A/E, 1906–1920	0.053	3.65	0.0003	0.0005
A/E, 1921–1926	0.046	4.71	<0.0001	<0.0001

Panel A: restricted wild-cluster bootstrap p -values (Rademacher weights, 9,999 draws, clusters are insurers) after partialling out the size cubic and region-by-year effects, for the New York-sample estimates of Table IA.14, Panel A (own death claims from the New York reports and the Spectator’s claims on the same firm-years, in percent of the mean), and of Table IA.20a, Panel A, columns 3 and 4; the last column is the number of insurers. Panel B: conventional p -values and Benjamini–Hochberg false-discovery-rate q -values for the nine estimates of Table IA.13, Panel A, row 1, Panel B, row 1, and Panel C, row 1; the last column is the q -value.

Table IA.22: Survival, exit channels, and mediation through size
IA.22a. Stability: competing exits

	Survival (10 yr)	Failure (10 yr)	Consolidation (10 yr)	Retirement (10 yr)	LR volatility (5 yr)
Breadth, unconditional	0.077*** (0.012)	-0.004 (0.005)	-0.066*** (0.011)	-0.007** (0.003)	-0.016*** (0.001)
Breadth, conditional on size (Table XI)	0.001 (0.016)	0.000 (0.007)	0.003 (0.015)	-0.004** (0.002)	0.001 (0.001)
Breadth _{t-5} size _{t-5} : total	-0.016 (0.017)	0.000 (0.007)	0.017 (0.017)		-0.000 (0.002)
+ size _t : direct	-0.025 (0.018)	0.000 (0.006)	0.026 (0.017)		0.002 (0.002)
Mean of outcome	0.779	0.018	0.199	0.004	0.068

Firm-year regressions with region-by-year effects; standard errors clustered by firm. Failure, consolidation (reinsured, merged, acquired), and retirement within ten years sum to one minus survival. The last two rows measure breadth and size five years before the outcome's start year; "direct" adds a cubic in current size.

Table IA.22: Survival, exit channels, and mediation through size (continued)
IA.22b. Mediation through size

	$h = 5$		$h = 10$	
	Region×year FE	+ firm FE	Region×year FE	+ firm FE
Size _t on breadth _{t-h} (first stage)	0.052*** (0.012)	0.043* (0.024)	0.046** (0.020)	0.001 (0.026)
Ten-year survival: total	-0.016 (0.017)	-0.025 (0.020)	-0.035* (0.020)	-0.036 (0.027)
direct (adding size _t)	-0.025 (0.018)	-0.028 (0.020)	-0.042** (0.021)	-0.036 (0.027)
Ten-year consolidation: total	0.017 (0.017)	0.035** (0.018)	0.039** (0.020)	0.055** (0.025)
direct (adding size _t)	0.026 (0.017)	0.037** (0.018)	0.046** (0.020)	0.055** (0.025)
Loss-ratio volatility: total	-0.000 (0.002)	-0.001 (0.002)	0.001 (0.002)	0.001 (0.002)
direct (adding size _t)	0.002 (0.002)	0.000 (0.002)	0.003 (0.002)	0.001 (0.002)
Firm-years (first stage)	6,782	6,735	4,842	4,804

Breadth and size are measured h years before the outcome's start year, with a cubic in size at $t - h$. The first row regresses size at t on breadth at $t - h$; "total" rows regress the outcome on breadth at $t - h$; "direct" rows add a cubic in size at t . Region-by-year effects, and firm effects where indicated; standard errors clustered by firm.

Table IA.23: Claim-count noise
IA.23a. Claim-count noise and the source of average policy size

<i>Panel A. Breadth coefficient, conditional on a size cubic, by source of average policy size</i>					
Policy size	$\ln N$	$\ln(R - N)$	N/R	$\operatorname{asinh} \frac{(R - N)}{\tilde{R}}$	Windows (share $R \leq N$)
Mixed sources (Table XII)	0.140** (0.061)	0.055 (0.068)	0.011 (0.016)	-0.060 (0.045)	7,271 (0.17)
NY median, every firm	0.126*** (0.032)	0.078 (0.066)	0.016 (0.016)	-0.055 (0.045)	7,271 (0.18)
Firm's own NY statement	-0.156 (0.147)	-0.093 (0.172)	0.017 (0.023)	-0.082 (0.116)	1,660 (0.05)
<i>Panel B. Components of $\ln N$: loss ratio, policy size F, premium, mixed sources</i>					
	$\ln N$	$\ln$ loss ratio	$\ln F$		
Breadth	0.140** (0.061)	0.144*** (0.029)	0.013 (0.049)		

Windows of Table XII; windows with a nonpositive realized variance or claim-count component and windows alone in their region-year are dropped, leaving 7,271 of the 7,285 in the regressions. N is recomputed with the average policy size of Table XII (mixed sources), with the New York median policy size of the year for every firm, or with the firm's own New York in-force average where it is observed in all five years. The asinh column is the inverse hyperbolic sine of $(R - N)$ divided by the median R , which keeps windows with $R \leq N$; the last column reports the number of windows and, in parentheses, the share with $R \leq N$. Panel B regresses the log of N , of the window's mean loss ratio, and of its mean policy size on breadth. Breadth and size are standardized logs of the window means; region-by-year-of-window effects; firm clusters. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table IA.23: Claim-count noise (continued)
IA.23b. Claim-count noise with each insurer's own policy size and number of deaths

Claim-count noise N	$\ln N$	$\ln(R - N)$	N/R	$\operatorname{asinh} \frac{(R - N)}{\tilde{R}}$	Windows (share $R \leq N$)
Mixed sources (Table XII), all windows	0.140** (0.061)	0.055 (0.068)	0.011 (0.016)	-0.060 (0.045)	7,271 (0.17)
Own in-force policy size where reported, all windows	0.154** (0.061)	0.061 (0.067)	0.017 (0.015)	-0.063 (0.045)	7,271 (0.15)
Mixed sources, windows with death counts	0.213*** (0.061)	0.226* (0.121)	-0.045 (0.030)	0.157* (0.084)	1,463 (0.14)
Death counts in place of F , same windows	0.217*** (0.062)	0.201 (0.127)	-0.039* (0.023)	0.144* (0.077)	1,463 (0.10)

Windows of Table IA.23a. Row 1 reproduces its first row. Row 2 replaces the average policy size F with the insurer's own average policy in force from the Spectator's company accounts where reported (between \$200 and \$20,000), else keeps the F of Table XII. Row 4 replaces the claim-count component with the Poisson variance of the loss ratio implied by the insurer's number of deaths in the year, $(\text{losses/premium})^2/\text{deaths}$, averaged over the window, on the windows in which deaths are reported in all five years; row 3 is row 1 on those windows. Insurers with industrial business keep the F of Table XII and have no death counts. The last column reports the number of windows and, in parentheses, the share with $R \leq N$. Breadth and size are standardized logs of the window means; region-by-year-of-window effects; firm clusters. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table IA.24: A geographic benchmark for the stability test
IA.24a. A geographic benchmark for the stability test

	Estimate	Windows / test
<i>Panel A. Five-year windows on breadth (size cubic, region-by-year effects)</i>		
$\ln G$, geographic variance	-0.715*** (0.059)	7,271
$\ln(R - N)$, variance beyond claim-count noise	0.055 (0.068)	6,056
$\operatorname{asinh}((R - N)/\tilde{R})$	-0.060 (0.045)	7,271
<i>Panel B. Share of G in $R - N$ (median)</i>		
All windows (mean in brackets)	0.106 [0.290]	6,065
Smallest third by size	0.143	2,022
Middle third	0.145	2,022
Largest third	0.053	2,021
<i>Panel C. Equivalence test for $\ln(R - N)$</i>		
Estimate (90% confidence interval)	0.055	[-0.057, 0.166]
Benchmark: full geographic diversification, median share	-0.056	$p = 0.051$
Benchmark: full geographic diversification, mean share	-0.160	$p = 0.001$

Windows of Table IA.23a. The geographic shock of a market cell is the loss ratio of the other insurers in the state and year less the national loss ratio of the year; a firm-year's shock is the premium-weighted sum over its cells, and G is its variance over the window. Panel A regresses $\ln G$, $\ln(R - N)$, and the inverse hyperbolic sine of $(R - N)$ over the median R on breadth, with a size cubic, region-by-year-of-window effects, and firm clusters. Panel B: $G/(R - N)$, capped at 1, on windows where both logs are defined, overall and by thirds of size. Panel C: the benchmark is the effect on $\ln(R - N)$ if breadth reduced the geographic part of $R - N$ as it reduces G , $\ln(1 - s + s e^{b_G})$, where s is the median or mean share and b_G the coefficient on $\ln G$; p is the one-sided p -value of the test that the coefficient on $\ln(R - N)$ is no greater than the benchmark. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table IA.25: Common risk across markets and the 1918 pandemic
IA.25a. The share of firm loss-ratio variance common across markets

Breadth quartile	Q1	Q2	Q3	Q4
Median effective number of states	1.24	2.32	5.62	12.26
<i>Panel A. Benchmark built from state volatilities</i>				
Systematic share, all cells	0.89	0.92	0.92	0.95
median of window-level $1 - b^2/r^2$, all cells	0.18	0.50	0.66	0.73
<i>Screened cells:</i>				
Mean realized volatility	0.092	0.088	0.060	0.048
Mean independence benchmark	0.056	0.041	0.028	0.019
Systematic share, $1 - (\bar{b}/\bar{r})^2$	0.62	0.78	0.79	0.85
excluding top 5% most volatile windows	0.38	0.61	0.70	0.78
median of window-level $1 - b^2/r^2$	0.28	0.53	0.66	0.73
<i>Panel B. Benchmark built from the firm's own state loss ratios</i>				
Cross-market share, median $1 - B/R$	0.09	0.24	0.30	0.44
stable books only	0.04	0.12	0.24	0.43
excluding 1918–19 and 1930–33	0.06	0.20	0.18	0.31
Share of windows with $B < R$	0.63	0.69	0.66	0.71
<i>Panel C. Claim-count sampling noise</i>				
Median noise variance / realized variance	0.37	0.19	0.10	0.06
windows with policy counts	52	172	581	946
Firm windows	1408	1408	1408	1407

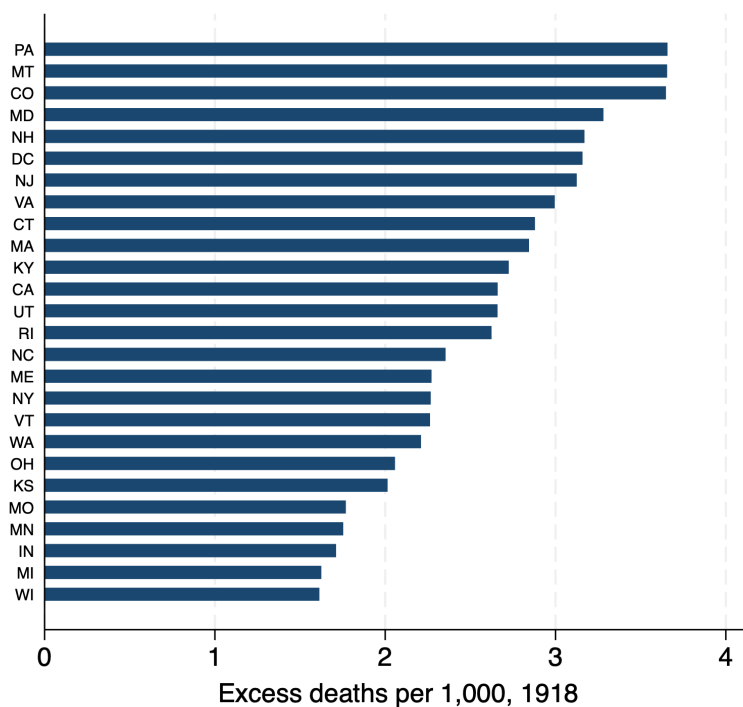
Five-year windows of consecutive firm-years for firms writing in at least two states, as in Figure IX. r and R are the standard deviation and variance of the firm's loss ratio in the window; b is a benchmark $(\sum_s w_{is}^2 v_s^2)^{1/2}$ built from the volatility v_s of each state's aggregate loss ratio; $B = \sum_s \bar{w}_{is}^2 \text{Var}(lr_{ist})$ over states written in all five years. Except where marked, cells failing the validity screen are excluded. Claim-count noise is losses $\times F/\text{premium}^2$ with F the firm's average new policy size, where policy counts are reported.

Table IA.25: Common risk across markets and the 1918 pandemic (continued)
IA.25b. The 1918–1919 influenza: power of the firm-level test

	Loss-ratio change on exposure	MDE	S.d. of exposure	Implied by market slope	S.d. of loss- ratio change	Firms
All insurers	0.010 (0.037)	0.103	0.549	0.0101	0.125	133
Claim-count noise below 25% of variance	0.029 (0.045)	0.125	0.370	0.0068	0.082	52
Claim-count noise below 10% of variance	-0.073 (0.065)	0.183	0.288	0.0053	0.078	28

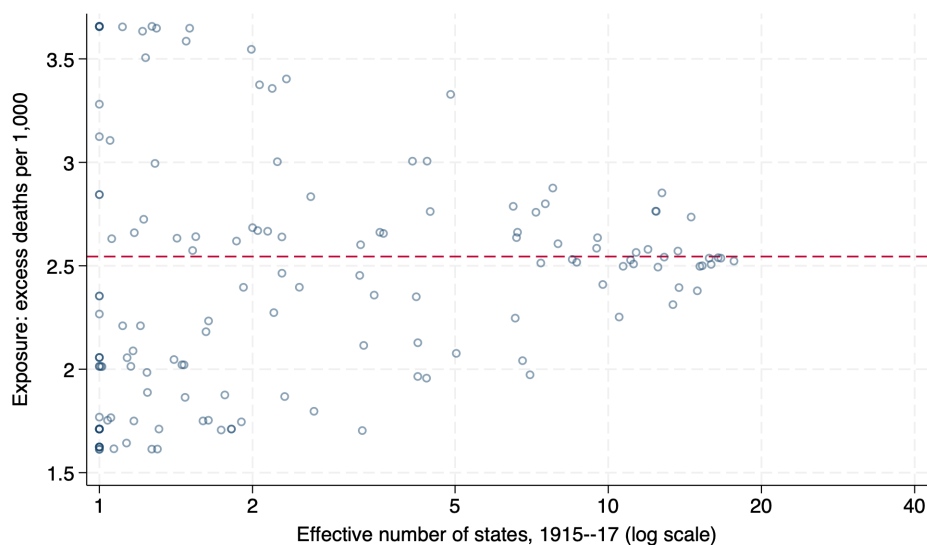
The firms of Table IA.26a, Panel E: the change in each firm's loss ratio from 1915–1917 to 1918–1919 regressed on its premium-weighted exposure to state excess mortality (deaths per 1,000) and on size, with region fixed effects and robust standard errors. MDE is the minimum detectable effect at the 5 percent level with 80 percent power, 2.8 times the standard error. *Implied by market slope* is the market-level coefficient of Table IA.26a, Panel E, multiplied by the standard deviation of exposure. Claim-count noise is the median share of the firm's loss-ratio variance due to claim counts in five-year windows starting 1911–1913 (Table XII). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure IA.8: Excess mortality in the 1918 influenza pandemic, by state



Notes. Excess deaths per 1,000 inhabitants in 1918, states of the death registration area, ordered from the most to the least affected. Source: U.S. Bureau of the Census (1920).

Figure IA.9: Breadth and exposure to the 1918 pandemic



Notes. Each point is an insurer of the sample of Table XIII. The horizontal axis is the insurer's effective number of states in 1915–1917, on a log scale; the vertical axis is its exposure, the premium-weighted average of 1918 excess deaths per 1,000 across the states in which it wrote. The dashed line is the premium-weighted mean exposure.

Table IA.26: Secondary robustness
IA.26a. Secondary robustness

<i>Panel A. Survival (Table II): breadth coefficient</i>				
	Unconditional	Region $\times$ year FE Size cubic	HQ state $\times$ year FE Unconditional	Size cubic
5-year survival	0.052*** (0.008)	-0.001 (0.010)	0.053*** (0.009)	-0.001 (0.012)
10-year survival	0.077*** (0.012)	0.001 (0.016)	0.077*** (0.013)	-0.004 (0.018)
<i>Panel B. Within-market level regression (Table IX), clustered by firm and state-year</i>				
	Breadth	$\times$ local share	$\times$ tenure (10 yrs)	
Loss ratio	0.026*** (0.008)	0.172 (0.110)	-0.004 (0.004)	
Claims per \$1,000	0.74*** (0.21)	3.99 (3.80)	-0.01 (0.11)	
Premium per \$1,000	-0.47* (0.24)	-18.88*** (4.33)	0.50*** (0.18)	
	Breadth	$\times$ HQ-state market	$\times$ largest market (non-HQ)	
Loss ratio	0.033*** (0.008)	-0.020*** (0.005)	-0.059*** (0.009)	
Claims per \$1,000	0.88*** (0.21)	-0.61*** (0.15)	-1.88*** (0.29)	
Premium per \$1,000	-0.43* (0.25)	-0.41** (0.21)	-1.03*** (0.32)	
<i>Panel C. Distance gradient (Table VI), clustered by firm and state-year</i>				
	State centroid + tenure		Largest city + tenure	
Loss ratio	-0.039*** (0.004)	-0.029*** (0.003)	-0.033*** (0.003)	-0.024*** (0.003)
Premium per \$1,000	-0.739*** (0.087)	-0.583*** (0.089)	-0.625*** (0.076)	-0.484*** (0.078)
<i>Panel D. Section 6 medians by breadth quartile, firm-block bootstrap s.e. (500 draws)</i>				
	Q1 (narrow)	Q2	Q3	Q4 (broad)
Covariance share, $1 - B/R$	0.09*** (0.02)	0.24*** (0.03)	0.30*** (0.04)	0.44*** (0.04)
Claim-count noise share	0.37*** (0.12)	0.19** (0.09)	0.10*** (0.04)	0.06*** (0.01)
Q4 - Q1 covariance share	0.35*** (0.04)			
<i>Panel E. The 1918-19 influenza (change in loss ratio, 1918-19 vs 1915-17)</i>				
Market: on state excess deaths per 1,000	0.018** (0.009)			
Firm: on premium-weighted excess deaths	0.010 (0.037)			
Firm: exposure deviation on breadth	-0.147*** (0.051)			
Firm: loss-ratio shock deviation on breadth	0.003 (0.012)			

Panel A repeats Table XI, Panel A, with headquarters-state-by-year effects. Panel B repeats the Table IX specification with local share and tenure, adding interactions of breadth with the centered local share and with tenure in decades (upper block), and with indicators for the firm's headquarters state and for its largest market when that market is not its headquarters state (lower block); two-way clustered standard errors. Panel C repeats Table VI, columns 1, 2, 5, and 6, with distance measured to the state's centroid and to its largest city. Panel D reports the medians of Section 6 with standard errors from 500 firm-block bootstrap draws, which allow for the overlap of adjacent windows. Panel E compares loss ratios in 1918-1919 with 1915-1917: the market row regresses the change in a firm-state loss ratio on the state's excess deaths per 1,000 in the pandemic, weighted by 1915-1917 premium and clustered by state; the firm rows use firms with at least half of their 1915-1917 premium in states with excess-mortality data, with exposure measured as the premium-weighted excess mortality of those states, region effects, and robust standard errors.