Received: 14 May 2019 | Revised: 26 October 2020 | Accepted: 22 November 2020

DOI: 10.1111/jori.12336

ORIGINAL ARTICLE

The Journal of Risk and Insurance

Estimating extreme cancellation rates in life insurance

Francesca Biagini[1] | **Tobias Huber**[2] | **Johannes G. Jaspersen**[3] | **Andrea Mazzon**[1]

[1]Workgroup Financial and Insurance Mathematics, Munich Risk and Insurance Center, Ludwig-Maximilians-Universität München, Munich, Germany

[2]Institute for Risk Management and Insurance, Munich Risk and Insurance Center, Ludwig-Maximilians-Universität München, Munich, Germany

[3]Institut für Versicherungsbetriebslehre, Gottfried Wilhelm Leibniz Universität Hannover, Hanover, Germany

Correspondence
Tobias Huber, Institute for Risk Management and Insurance, Munich Risk and Insurance Center, Ludwig-Maximilians-Universität München, Geschwister-Scholl-Platz 1, 80539 Munich, Germany.
Email: tobias.huber@lmu.de

Abstract

This paper assesses the risk of a mass lapse event in life insurance. The rarity of the event and the complexity of policyholder behavior make the risk assessment of such a scenario difficult. Using a simulation study, we evaluate how different estimation methods can assess the risk of this scenario, using panel data at the company level. We then use the best-performing method to estimate the probability distribution function of a mass cancellation event in the United States and Germany. We identify dependencies of the event on company and country characteristics, which have not been taken into account by regulating agencies. We also find that the current mass lapse scenario in Solvency II has no empirical foundation for the German market. We show that an empirically valid scenario leads to a significantly lower solvency capital requirement for the average German life insurer.

KEYWORDS

dynamic peaks over threshold, extreme value theory, life insurance, mass cancellation

JEL CLASSIFICATION

C14; G18; G22; G32

1 | INTRODUCTION

Rare events with extreme consequences often have a profound influence on economies and their economic agents. The 9/11 terrorist attacks led to considerable changes in many economic sectors, such as the airline industry. Hurricanes, such as Katrina and Sandy, have shaped how we see flood protection and insurance. Financial crises, such as the Great Depression and that of 2009, influence our considerations of financial regulation and economic risks, in general. For organizations, extreme events can similarly shape existence. One such extreme event for life insurance companies is the occurrence of a mass cancellation event, that is, a large portion or even a majority of the policyholders who cancel their life insurance policy abruptly. Even though policyholder-cancellation behavior has received considerable attention in the academic literature (e.g., Eling & Kiesenbauer, 2013; Kuo et al., 2003), the question of how to model mass cancellation scenarios has received scant attention. This is surprising because the possibility of mass cancellations has a large effect on insurance companies' asset liability management and leads to one of the largest financial reserves in the European risk management framework Solvency II (EIOPA, 2011).

This omission in the literature can be explained, in part, by the data requirements associated with estimating extreme cancellation events. Extreme cancellation, and its associated rates, is a rare event; thus, data sourced from only one insurer are not sufficient to estimate this tail risk. In this paper, we address this challenge by employing the estimation approach of Chavez-Demoulin et al. (2016), which is based on extreme value theory and can be applied to panel data. We consider cancellation rates as realizations of random variables but do not make any assumptions about their distribution functions, except that they are continuous.[1] This method enables us to estimate the unknown probability distribution functions and associated risk measures for extreme cancellation events from readily available panel data at the company level.

Dependent on contract characteristics, cancellations can either negatively or positively affect an insurer's profits. Unexpected changes in the level of cancellation rates can lead to liquidity problems, the loss of expected future profits, and unbalanced initial expenses (Eling & Kiesenbauer, 2013; Kuo et al., 2003). Life insurers are, thus, interested in assessing their exposure to this risk as accurately as possible. U.S. life insurers can potentially face very high cancellation rates, particularly after premium guarantee periods expire (SOA & LIMRA, 2018, p. 29). The possibility of such an extreme event needs to be taken into consideration for a company's asset liability management. In other situations, however, cancellations can increase an insurer's profits, as policies are usually front-loaded, and a cancellation allows insurers to reap the early profits without having to incur expected losses in the later part of the policy's life cycle (Gottlieb & Smetters, 2020). In each case, understanding cancellation behavior is crucial for life insurers to ensure adequate asset liability management.

For the European market, the importance of understanding the cancellation behavior of policyholders has become critical with the introduction of the new regulatory framework Solvency II. The framework's solvency capital requirement (SCR) has a great sensitivity to lapse risk (EIOPA, 2011, p. 67). In the standard model of the regulation framework, the mass lapse

[1]The continuity assumption enables us to obtain results on the far end tail of the cancellation rates' distribution functions under technically mild assumptions. The literature shows that the conditions for the existence of a limit distribution for discrete distribution functions are strong (Davison & Huser, 2015; Hitz et al., 2017; Nadarajah & Mitov, 2002). Our continuity assumption allows us to consider a more flexible framework. The following reasons justify our continuity assumption: (1) We observe no evidence for a discontinuous distribution function in the histograms of cancellation rates stemming from two different countries (United States and Germany) between 1996 and 2018 (see Figures 1 and 4). (2) The commonly assumed sources (e.g., unemployment rate or interest rate) of cancellation rates are modeled as continuous random variables.

shock leads companies to have solvency capital requirements in the hundreds of millions (Old Mutual, 2016; UNIQA, 2017) because insurers have to apply a scenario in which 40% of all policyholders cancel their contract in the same year. This assumption does not have an empirical justification for the assumed size of the shock or account for the fact that the same shock is assumed for all of the different national markets in Europe. Our model offers a way to use empirical assumptions appropriate for the individual national insurance market in the lapse-risk module, which can help to limit the risk of possible over- or under-reserving. Previous literature on cancellation events, in general, has used two approaches—estimating cancellation rates from data directly (e.g., Eling & Kiesenbauer, 2013; Fang & Kung, 2020; Knoller et al., 2016) or calibrating economic models based on specific assumptions about macroeconomic factors and policyholder behavior (Albizzati & Geman, 1994; Bacinello, 2003; Consiglio & De Giovanni, 2010). These approaches differ in their interpretation. Empirical approaches often make fewer assumptions than do other approaches but seldom allow for a causal interpretation of the results. Model-based approaches identify causal relationships but assume specific relationships between policyholder behavior and macroeconomic factors, which may not hold. For extreme cancellation events, some theoretical modeling approaches exist (Barsotti et al., 2016; Loisel & Milhaud, 2011), but, as of yet, no empirical estimations have been reported.

Applying the estimation method to data sourced from U.S. life insurers demonstrates the applicability of our approach and shows the importance of the product type for high cancellation rates in the U.S. market. Although mass cancellation scenarios with rates up to 50% can be adequate for companies with a high share of term life policies, a mass cancellation scenario for companies with a high share of permanent life insurance policies is estimated to be approximately 20%. Further, we consider the calibration of Solvency II's mass cancellation scenario, using German data. Our results show that cancellation rates of 20%–25% reflect a mass cancellation scenario in the German life insurance market. These values raise skepticism as to whether the arbitrarily chosen cancellation rate of 40% for this scenario in Solvency II is adequate for the German life insurance market. In both the U.S. and German markets, the severity of the extreme cancellation event is decreasing in correspondence with the size of the insurers' portfolio and increasing in proportion to the amount of new business. This encourages further research on other company-specific variables. The combination of these results and that the mass cancellation scenario can depend on the predominant product type sold in a market call into question whether a uniform mass cancellation scenario for all European life insurance markets is appropriate. European life insurance markets differ in their company characteristics (Insurance Europe, 2019) and in their dominant products (Standard & Poors, 2018) and might, thus, differ in their mass cancellation scenario as well.

We view the contribution of our paper as complementary to, rather than competing with, model-based approaches. Although our approach does not rely on assumptions about policyholder behavior, we are able to estimate only a cancellation rate distribution but are unable to identify the causal relationship that leads to the cancellation rate. Model-based approaches can use our results to validate their behavioral assumption by observing whether their models produce similar cancellation rate distributions. Alternatively, our results can be used directly for models of cancellation behavior.

In the following, we first summarize the empirical estimation method and provide a simulation study to examine the approach in a panel data context. We then apply the estimation procedure to U.S. and German data. The next section provides implications for the modeling of extreme cancellation rates and for insurance regulation. The paper ends with concluding remarks and provides directions for future research.

2 | METHODOLOGY

2.1 | Overview

Our approach is based on considering cancellation rates as a stochastic variable. The main hypotheses that prevail in the literature to explain cancellation behavior are based on macroeconomic variables, such as the interest or unemployment rate.[2] As macroeconomic variables can be seen as stochastic processes, and changes in macroeconomic variables lead to changes in cancellation rates, cancellation rates also can be modeled stochastically. A major problem with assessing the underlying distribution function, however, is posed by the lack of sufficient extreme data. We thus employ the peaks-over-threshold (POT) method, which exploits the existence of a natural candidate for the conditional distribution function above a high level.

In this section, we first summarize the POT method as it is described in Embrechts et al. (1997) and McNeil et al. (2015). In its original form, the method can be applied only to independent and identically distributed (i.i.d.) data. For our purposes, this would require a long time series of cancellation rates for a single company. Such data, however, are not reliably available. We are limited to using panel data of multiple companies over short periods of time. These data feature a dependence structure along two dimensions. First, macroeconomic variables affect all companies at the same time so that it can be assumed that observed cancellation rates within the same time period are correlated. Second, companies can have idiosyncratic factors that lead to higher or lower cancellation rates, which will make the observations within a company serially correlated over time. To address these issues, we summarize a dynamic version of the POT method, which was introduced by Coles (2001) for parametric dependencies and extended by Chavez-Demoulin et al. (2016) to semiparametric and nonparametric dependencies. The dynamic POT method takes into account the dependency structure of our panel data by modeling the parameters of the estimated probability distribution dependent on time and on a quantitative company-level variable (in our case, the portfolio size). Finally, this section provides a simulation study to evaluate the dynamic POT method in a panel data context.

2.2 | Peaks over threshold method

We consider a panel of n insurance companies over T periods with cancellation rates $y_{j,s}$, modeled as realizations of the random variables $Y_{j,s}(j = 1, 2, ..., n; s = 1, 2, ..., T)$. We first make the strong assumption that all random variables $Y_{j,s}$ are i.i.d. with cumulative distribution function F. The classic POT method then allows us to reliably estimate the risk of extreme cancellation rates without knowing the common underlying distribution function F.

The intuition is to split the distribution function F into two parts based on a "suitably large" threshold $u > 0$. Below the threshold, we are given enough data so that the empirical distribution provides a good fit. Above the threshold, the Pickands-Balkema-de Haan Theorem (e.g., Balkema & de Haan, 1974; Pickands, 1975) enables us to approximate the excess distribution function of F over u by the generalized Pareto distribution (GPD) function.

[2]This is mostly based on two concepts: the interest rate hypothesis and the emergency fund hypothesis. Evidence for the former can be found in Schott (1971), Pesando (1974), and Kuo et al. (2003). Evidence for the latter is reported in Dar and Dodds (1989), Outreville (1990), and Kiesenbauer (2012).

We define the excess distribution function F_u of $Y_{1,1}$ (all $Y_{j,s}$ are i.i.d.) over a threshold $u, 0 < u < z_F := \sup\{z \in \mathbb{R}: F(z) < 1\}$, by the conditional probability:

$$F_u(z) := \mathbb{P}(Y_{1,1} - u \le z \,|\, Y_{1,1} > u) = \frac{F(z+u) - F(u)}{1 - F(u)}, \quad 0 \le z \le z_F - u. \tag{1}$$

Rearranging this equation and defining $\bar{F} := 1 - F$ and $\bar{F}_u := 1 - F_u$, we obtain a method for estimating the far end tail of F by estimating $\bar{F}(u)$ and $\bar{F}_u(z)$:

$$\bar{F}(z+u) = \bar{F}(u) \cdot \bar{F}_u(z). \tag{2}$$

We approximate $\bar{F}(u)$ by the empirical distribution function $\bar{F}_N(u) := 1 - F_N(u)$ and approximate $\bar{F}_u(z)$, using a GPD.[3] This distribution function is dependent on a shape parameter $\xi \in \mathbb{R}$ and a scale parameter $\beta > 0$ and is given by

$$G_{\xi,\beta}(z) = \begin{cases} 1 - \left(1 + \frac{\xi z}{\beta}\right)^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp\left(-\frac{z}{\beta}\right), & \xi = 0, \end{cases} \tag{3}$$

where $z \ge 0$ if $\xi \ge 0$ and $z \in [0, -\beta/\xi]$ if $\xi < 0$. Because the realizations $y_{j,s}$ are i.i.d., $\bar{F}_u$ can be approximated by $\bar{G}_{\hat{\xi},\hat{\beta}(u)} := 1 - G_{\hat{\xi},\hat{\beta}(u)}$ with estimates $\hat{\xi}$ and $\hat{\beta}(u)$ of the threshold-independent shape parameter ξ and the threshold-dependent scale parameter $\beta(u)$ (McNeil et al., 2015).

The estimates $\hat{\xi}$ and $\hat{\beta} = \hat{\beta}(u)$ can be computed by the method of maximum likelihood applied to the set of data y_{j,s_j} with $s_j \in S_j$, where $S_j \subseteq \{1, 2, ..., T\}$ denotes the subset of points in time at which the excesses over the threshold u for company j occur. In addition, a maximum likelihood estimate of $\bar{F}_N(u)$ is given by N_u/N, where N_u denotes the total number of excesses, and $N := nT$ designates the total number of cancellation rates. Given these estimates and a threshold u, the estimated quantile $\hat{q}(\alpha)$ of the unknown distribution function F at confidence level α is equal to:

$$\hat{q}(\alpha) = u + \frac{\hat{\beta}}{\hat{\xi}}\left[\left(\frac{1-\alpha}{N_u/N}\right)^{-\hat{\xi}} - 1\right]. \tag{4}$$

The choice of the threshold u is an important component of the POT method: If the threshold is too low, the exceedances also include nonextreme events, and the estimates are biased. In contrast, if the threshold is too high, the number of exceedances is low, and the variance in the statistical estimation is large. To select a threshold, the goodness of fit of empirical excesses over the chosen threshold u to a parametric GPD model can, for example, be evaluated through a Q-Q plot. In such a graph, the quantiles of the log-transformed excesses over u are plotted against the theoretical quantiles of an exponential distribution. If a straight line is observed, then it is empirically confirmed that the GPD provides a good fit of the data. The optimal threshold is chosen as the smallest u for which a good fit is observed.

For commonly available panel data, the assumptions of i.i.d. random variables are usually violated. Thus, in our later applications, we employ a generalization of the POT method for

[3] The empirical distribution function F_N of F given $N = nT$ i.i.d. observations of n companies over T periods is defined as: $F_N(z) := \frac{1}{N}\sum_{j=1}^{n}\sum_{s=1}^{T} \mathbb{1}_{(y_{j,s} \le z)}, z \ge 0$.

non-i.i.d. data, which enables us to let the frequency and severity of the excesses be dependent on covariates.

2.3 | Dynamic peaks over threshold method

Again, ξ and β are parameters of the GPD that model the size of the excesses over a large threshold. The number of excesses is assumed to follow a nonhomogeneous Poisson process with rate function λ. The dynamic POT method allows the frequency parameter λ and the severity parameters ξ and β to be dependent on covariates, which avoids making the assumption that the cancellation rates are realizations of i.i.d. random variables (Chavez-Demoulin et al., 2016; Coles, 2001). In our applications, we use two covariates: the first covariate x represents the insurer's portfolio size, and the second covariate s represents time. The reparameterization $\nu := \log(\beta(1+\xi))$ and general measurable functions $g_\xi, g_\nu, g_\lambda \colon \mathbb{R}_{\geq 0} \to \mathbb{R}$ and $h_\xi, h_\nu, h_\lambda \colon [1, T] \to \mathbb{R}$ leads to the following generalized additive models for the parameters of the dynamic POT method:[4]

$$\xi = \xi(x, s) = g_\xi(x) + h_\xi(s), \tag{5}$$

$$\nu = \nu(x, s) = g_\nu(x) + h_\nu(s), \tag{6}$$

$$\lambda = \lambda(x, s) = g_\lambda(x) + h_\lambda(s). \tag{7}$$

In the above equations, g_ξ, g_ν, g_λ and h_ξ, h_ν, h_λ are either linear or smooth functions. In their combination they can model a parametric, semi-parametric, or nonparametric dependence of the parameters on the corresponding covariates. We utilize natural cubic splines for the smooth functions and determine their degrees of freedom and, thus, their smoothness based on Akaike's information criterion (AIC) value (see Section 3.1.3 and Chavez-Demoulin et al. (2016, p. 746) for more details).

Pooling the N_u excesses y_{j,s_j} under the consideration of the corresponding portfolio size covariate $x = x_{j,s_j}$ and time covariate $s = s_j$ enables us to obtain estimates for $\xi(x, s)$ and $\nu(x, s)$ by employing the penalized maximum likelihood estimation of Chavez-Demoulin et al. (2016). Moreover, we can directly estimate the rate $\rho(x, s) := \lambda(x, s)/N(x, s)$ of the nonhomogeneous Poisson process through a logistic regression model, where $N(x, s)$ stands for the total number of observations for a fixed covariate x and time point s. Subsequently, the estimated α-quantile $\widehat{q}(\alpha)$ for a threshold u in dependence of the covariates x and s is given by

$$\widehat{q}(\alpha)(x, s) = u + \frac{\widehat{\beta}(x, s)}{\widehat{\xi}(x, s)}\left[\left(\frac{1-\alpha}{\widehat{\rho}(x, s)}\right)^{-\widehat{\xi}(x,s)} - 1\right]. \tag{8}$$

The applicability of the dynamic POT method has, for example, been shown by Chavez-Demoulin et al. (2016), Embrechts et al. (2018), Hambuckers, Groll, et al. (2018), and Hambuckers, Kneib, et al. (2018). Each of these papers applies the dynamic POT method to operational losses and uses covariates to control for the underlying heterogeneity across time and companies. As background to applying the method to estimate extreme cancellation rates, in the next section, we provide a simulation study to examine the dynamic POT method in a panel data context.

[4]The reparameterization $\nu := \log(\beta(1+\xi))$ guarantees the convergence of the simultaneous fitting procedure for the two parameters ξ and β (Chavez-Demoulin et al., 2016). This reparameterization requires $\xi > -1$, which is fulfilled in our applications.

2.4 | Simulation study

2.4.1 | Overview

We assess the suitability of the introduced methodology for applications to panel data. We aim to achieve two goals. First, we compare three different approaches with respect to their accuracy: In Case (1), we perform the classic POT estimation; in (2), we apply the dynamic POT to the companies' time series jointly, meaning that we take into account the panel structure of the data; and in (3), we employ the dynamic POT estimation to each company's time series separately. Second, we focus on threshold selection. Because we have heterogeneity in the parameters across time and within companies, we would optimally choose time-dependent and company-specific thresholds. Due to limited data availability, however, we are often restricted to choosing one threshold based on all of the pooled cancellation rate observations. We therefore examine the effect of such a simplification of the threshold selection on the estimation's accuracy in the dynamic POT method by comparing two approaches: In Case (2.1), we choose a joint empirical threshold based on pooling all data; and in (2.2), we choose an individual threshold for each company separately.

2.4.2 | Approach

We consider $n = 1000$ insurance companies over T years ($T = 20, 40, 80$) and simulate cancellation rates for each company $j = 1, 2, ..., n$ at time $s = 1, 2, ..., T$ according to a log-normal distribution $\mathcal{L}(\mu_{j,s}, \sigma_{j,s})$ with parameters $\mu_{j,s} \in \mathbb{R}$ and $\sigma_{j,s} > 0$. Each company j has at time s a company-specific covariate $x_{j,s} = x_{j,1}$, which varies from company to company but is constant over time. The parameters of the log-normal distribution are dependent on this quantitative covariate and are affected by a time-specific covariate $t_{j,s} = t_{1,s}$, which is constant across companies but varies over time. The covariates $x_{j,1}$ are chosen equidistantly between 0.2 and 0.4 for the first 900 companies. In addition, the covariate of company 901 is equal to 0.7 and increases in equidistant steps to 0.8 until company 1000. These 100 companies, thus, have markedly higher cancellation rates compared with the other companies. The time-specific covariate $t_{1,s}$ is equal to 0 at $s = 1$ and increases in equidistant steps to 0.1 until $s = T$. Overall, the parameters depend on the company- and time-specific covariates in the following way:[5]

$$\begin{aligned} \mu_{j,s} &= x_{j,1} + t_{1,s} - 3.5, \\ \sigma_{j,s} &= 1.1 \cdot x_{j,1} + 2 \cdot t_{1,s} + 0.1. \end{aligned} \tag{9}$$

Given the simulated cancellation rates and a threshold u, we apply the POT method and fit the excesses at time s_j over this threshold to a GPD. Because the cancellation rate observations follow a log-normal distribution with the parameters $\mu_{j,s}$ and $\sigma_{j,s}$, we can calculate the exact quantile $q_{j,s}(\alpha)$ for a chosen confidence level of $\alpha = 99.5\%$ by inverting the log-normal distribution. This enables us to define an error measure to evaluate the accuracy of estimating the α-quantile with the POT method. We run the simulation $R = 1,000$ times and obtain in each run $k = 1, 2, ..., R$ estimates $\hat{q}_{j,s_{j,k}}(\alpha)$ of the true quantile $q_{j,s_{j,k}}(\alpha)$ at the corresponding time

[5]The specifications in Equation (9) lead to summary statistics of the simulated cancellation rates ($Y_{j,s} \sim \mathcal{L}(\mu_{j,s}, \sigma_{j,s})$), which are similar to the cancellation rates observed in our empirical settings below.

$s_{j,k} \in S_{j,k} \subset \{1, 2, ..., T\}$ of the excess. We define the error measure as the mean relative difference between the estimated and true quantile and take the average over all R runs:

$$err = \frac{1}{R}\sum_{k=1}^{R}\left(\frac{1}{\sum_{i=1}^{n}|S_{i,k}|}\sum_{j=1}^{n}\sum_{s_{j,k}\in S_{j,k}}\frac{|\widehat{q}_{j,s_{j,k}}(0.995) - q_{j,s_{j,k}}(0.995)|}{q_{j,s_{j,k}}(0.995)}\right). \quad (10)$$

2.4.3 | Results

Table 1 provides the results of the error measure for the different estimation approaches (classic POT, simultaneous dynamic POT with a joint empirical threshold, simultaneous dynamic POT with an individual empirical threshold, one-by-one dynamic POT), and for different length of the time series ($T = 20, 40, 80$). All of these approaches are applicable only for a "suitably large" threshold. We do not have information on which threshold is high enough and, therefore, choose for all approaches four different thresholds equal to the 0%, 30%, 60%, and 90% quantiles of the

TABLE 1 Simulation study results

	20 years	40 years	80 years
(1) Classic POT			
0% quantile	0.310	0.309	0.309
30% quantile	0.859	0.854	0.853
60% quantile	0.926	0.924	0.923
90% quantile	0.773	0.774	0.772
(2) Dynamic POT: simultaneous estimation			
(2.1) Joint empirical threshold			
0% quantile	0.310	0.310	0.308
30% quantile	0.047	0.038	0.033
60% quantile	0.050	0.040	0.034
90% quantile	0.062	0.049	0.040
(2.2) Individual empirical threshold			
0% quantile	0.056	0.054	0.155
30% quantile	0.058	0.046	0.038
60% quantile	0.077	0.062	0.052
90% quantile	0.191	0.172	0.160
(3) Dynamic POT: one-by-one estimation			
0% quantile	0.550	0.467	0.494
30% quantile	–	0.390	0.230
60% quantile	–	–	0.259
90% quantile	–	–	–

Note: For different estimation approaches (classic POT, simultaneous dynamic POT, one-by-one dynamic POT) and for different lengths of the time series ($T = 20, 40, 80$), the error in estimating the 99.5% quantile of the underlying log-normal distribution as defined in (10) are displayed. For all approaches, we choose four different thresholds equal to 0%, 30%, 60%, and 90% quantiles of all cancellation rates. For the simultaneous dynamic POT, we also compare two threshold selection approaches, one that ignores and one that acknowledges heterogeneity in individual thresholds (joint empirical threshold, individual empirical threshold).

corresponding cancellation rates. For the classic POT or the dynamic POT with a joint empirical threshold, choosing the threshold equal to a specific quantile means that we have one threshold for all companies. In contrast, for the dynamic POT with an individual empirical threshold and for the one-by-one dynamic POT, this means that we have unique thresholds u_j for each company equal to the chosen quantile based on their corresponding cancellation rates. We thus compare for the simultaneous dynamic POT two threshold selection approaches, one that ignores and one that acknowledges heterogeneity in individual thresholds.

We begin by focusing on the results for a time series of 40 years (Table 1, Column 3). First, we compare the two versions of the simultaneous dynamic POT and find that the joint empirical threshold selection approach performs better than the individual empirical threshold approach in all cases, with the exception of the 0% quantile. The individual threshold is based on the empirical quantile of only 40 observations and is, therefore, affected by outliers. This makes the joint empirical threshold selection approach superior to the individual empirical threshold selection approach. Second, we compare the simultaneous dynamic POT with the joint empirical threshold to the one-by-one dynamic POT estimation and find marked differences in the estimation's accuracy. The one-by-one dynamic POT estimation is limited to data within a single time series, which allows a reliable estimation just for many data points. In the case of high thresholds, applying the dynamic POT estimation to each company's time series separately is not possible due to limited data availability. In cases for which the one-by-one dynamic POT estimation is applicable, this leads to a less-accurate estimation than does the simultaneous dynamic POT with the joint empirical threshold. Third, our results in Table 1 demonstrates that the simultaneous dynamic POT always performs better than the classic POT. The classic POT method has limited applicability in the case of data with heterogeneity across time and companies, which is commonly present in company-level panel data.

The results observed for a time series of 40 years are representative of those of time series with a different length (Table 1, Columns 2 and 4). In addition, we find that the estimation's accuracy improves with the length of the time period for all estimation approaches (with the exception of the 0% quantile). This effect is small for the classic POT estimation, considerably larger for the simultaneous dynamic POT estimation, and largest for the one-by-one dynamic POT estimation (see, e.g., the 30% quantile). In principle, the one-by-one dynamic POT estimation will likely become better than the simultaneous dynamic POT estimation if the number of time periods is sufficiently large. In applications, however, such long time series are usually not available, or at least not reliably available, due to changes in exogenous factors (e.g., regulation). In the range of time periods with reliable data, the results in Table 1 support the use of the simultaneous dynamic POT with the joint empirical threshold in our empirical settings, discussed below.

3 | EMPIRICAL ANALYSIS

3.1 | U.S. individual life insurance market

3.1.1 | Motivation

We employ the dynamic POT model to data from U.S. life insurers to demonstrate the applicability of our approach. Of all countries, the United States has the most readily available, reliable, and extensive data. Using only U.S. data does not allow us to make a general statement about life insurance markets but will give us a first indication of the effect of company-level covariates on extreme cancellation events. Moreover, it enables us to calibrate an empirically

justified mass cancellation scenario for the U.S. market. U.S. life insurers can potentially face very high cancellation rates, particularly after premium guarantee periods expire (SOA & LIMRA, 2018, p. 29). To analyze this risk, we use available panel data at the company level.

3.1.2 | Company-level, panel data

Our panel data include statistics on individual life insurance, collected from U.S. insurers' annual statement reports by the National Association of Insurance Commissioners (NAIC). The U.S. individual life insurance market distinguishes between two types of policies: Term insurance provides coverage for a predetermined period, whereas permanent insurance provides coverage for a whole life. Permanent insurance is either traditional whole life, universal life, variable life, or variable universal life insurance (ACLI, 2018). The data provide the direct written premiums as well as detailed portfolio information (e.g., policies in force, policies lost, policies issued) of 617 life insurers between 1996 and 2018. The reporting date of all the variables is the end of the year.

Although the data do not include cancellation rates directly, we are able to calculate cancellation rates for life insurer j in year s by dividing the number of policies that were canceled in s by the number of policies that were in force in s:

$$cancel_{j,s}^{US} := \frac{lap_{j,s}^{US} + sur_{j,s}^{US}}{lost_{j,s}^{US} + inForce_{j,s}^{US}} = \frac{lap_{j,s}^{US} + sur_{j,s}^{US}}{inForceProxy_{j,s}^{US}}. \tag{11}$$

In this definition lap^{US} denotes the contracts that were canceled as result of nonpayment of premiums, and sur^{US} designates contracts that were canceled and a cash surrender value paid to the policyholder. To capture all policies in force during the year ($inForceProxy^{US}$), we add all policies that were lost during the year ($lost^{US}$) to the policies in force at the end of the year ($inForce^{US}$).

We select our data to avoid reporting inconsistencies and to restrict the analysis to U.S. life insurance companies with a focus on primary insurance business. For each step of the data selection process, Table 2 provides the number of company-year observations and the number of companies kept with each data cleaning step.[6] The full sample includes 12,639 company-year observations of 617 life insurers between 1996 and 2018. In the first selection step, we keep companies that exhibit at least one positive direct written premium between 1996 and 2018. In the second and third steps, we keep companies with the share of policies assumed, which stem from reinsurance or coinsurance activities and never exceed 20% and keep companies that never exhibit a share of revived policies greater than 2%. These variables used in the sample selection process are given by the following ratios:

$$shareAssumed_{j,s}^{US} := assumed_{j,s}^{US} / inForceProxy_{j,s}^{US}, \tag{12}$$

$$shareRevived_{j,s}^{US} := revived_{j,s}^{US} / inForceProxy_{j,s}^{US}. \tag{13}$$

In the full sample, 97% of company-year observations exhibit a share of assumed policies smaller than 20%, and 97% of company-year observations exhibit a share of revived policies smaller than 2%. The first two steps of the data selection process, thus, restrict the analysis to active insurers with a focus on primary insurance business rather than a significant reinsurance

[6]Each permutation of the sample selection steps listed in Table 2 leads to the same selected sample.

TABLE 2 U.S. data: Sample selection process

Step	Description	Observations remaining	Companies remaining
0	All company-year observations between 1996 and 2018	12,639	617
1	Keep companies exhibiting at least one positive direct written premium between 1996 and 2018	12,010	574
2	Keep companies with $shareAssumed^{US}$ never exceeding 20%	8945	423
3	Keep companies with $shareRevived^{US}$ never exceeding 2%	7336	348
Selected sample		7336	348

Note: The steps of the sample selection process, the number of company-year observations, and the number of companies kept with each data-cleaning step are provided.

activity. The third step addresses the revival of lapsed policies. Because lapsed policies can be revived as long as 5 years after the lapse has happened, we are unable to identify whether a high number of lapses is the consequence of a high number of revivals in the subsequent year. The remaining sample after these three data selection steps represents 67% of aggregate direct written premiums and 35% of aggregate polices of the full sample.

Table 3 reports summary statistics of our company-level panel data in the selected sample. The reporting date of all the variables shown in the table is the end of the year. Direct written premiums are reported in US$ thousand, while variables on the portfolio are reported in number of policies. The descriptive summary of cancellation rates is in line with reported cancellation rates by SOA and LIMRA (2018) and ACLI (2018). We display the histogram of the cancellation rates in the

TABLE 3 U.S. data: Summary statistics

Variable	Description	*N*	Mean	Pctl(25)	Median	Pctl(75)
dwp^{US} ($000)	Direct written premium	7336	223,581	844	14,492	96,903
lap^{US}	Policies lapsed	7336	15,174	51	1075	6159
sur^{US}	Policies surrendered	7336	5742	29	647	3390
$lost^{US}$	Policies lost	7336	29,446	387	3592	16,189
$issued^{US}$	Policies issued	7336	24,671	15	1543	12,594
$assumed^{US}$	Policies assumed	7336	2952	0	0	0
$revived^{US}$	Policies revived	7336	692	0	16	281
$inForce_T^{US}$	Term life in force	7336	87,778	42	3129	32,616
$inForce_P^{US}$	Permanent life in force	7336	192,437	1888	28,051	106,109
$inForce^{US}$	Total in force	7336	290,120	3962	42,199	169,331
$shareTerm^{US}$	Share of term policies	7336	25.45%	0.74%	10.86%	39.89%
$shareIssued^{US}$	Share of issued policies	7336	9.05%	0.29%	4.95%	11.01%
$interest^{US}$	Interest rate changes	23	0.15pp	−0.35pp	0.11pp	0.43pp
$cancel^{US}$	Cancellation rates	7336	6.27%	2.89%	4.79%	7.23%

Note: Summary statistics of the variables used in the U.S. analysis of extreme cancellation rates are provided. The data include 348 life insurers between 1996 and 2018.

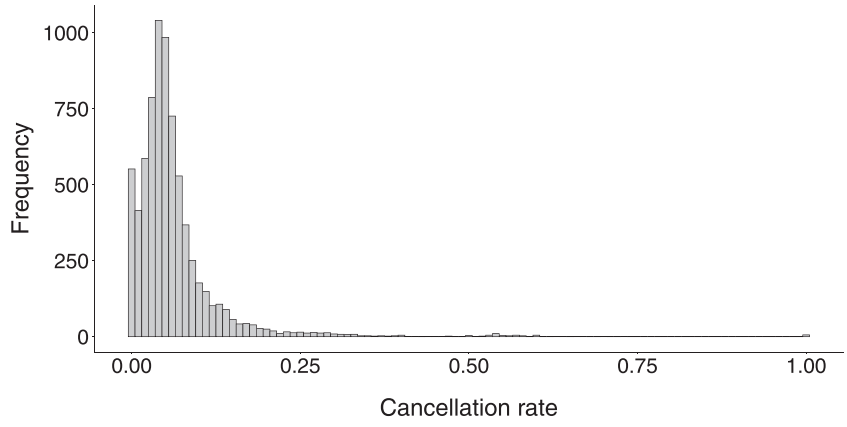

FIGURE 1 U.S. data: Histogram of cancellation rates. The distribution of the 7336 cancellation rates between 1996 and 2018 is displayed

selected sample in Figure 1. While the median of cancellation rates is equal to 4.79%, the histogram also shows cancellation rates above 30%. Because we are interested in a reliable estimate of the 99.5% quantile of the cancellation rates, employing the dynamic POT method seems particularly suitable given the few observations in the tail. To take the dependency structure of our panel data into account, we will use the number of policies in force ($x = inForceProxy^{US}$) and the year ($s = 1996, 1997, ..., 2018$) as covariates in the estimation. We also include the following control variables: (1) We monitor new business activity by the share of issued policies in terms of existing policies ($shareIssued^{US}$). (2) We track the focus of a life insurance company on term or permanent life insurance policies by the share of term life policies in force between 1996 and 2018 ($shareTerm^{US}$). (3) We control for changes in market interest rates by the first-difference of U.S. government bond yields with a duration of 10 years ($interest^{US}$).

3.1.3 | Model selection

The model specification consists of two steps: (1) threshold selection and (2) specification of the models for the frequency and severity of extreme cancellation rates. The specification is done via backward induction. For each threshold equal to the deciles of all cancellation rate observations, we perform model selection for the severity of extreme cancellation rates individually. Based on the chosen model for each threshold, we then select the lowest threshold with a good model fit according to a Q-Q plot. We describe the frequency and severity model selection process for the ex-post selected threshold of the 50% quantile of all cancellation rates, corresponding to $u = 4.79\%$ and 3668 excesses. We test whether the classic or dynamic POT method is adequate in our analysis based on likelihood ratio tests (LRTs) and the AIC value.[7] For the selection of the correct model specification within the dynamic POT method, we again base our selection on LRTs and AIC values. Whenever we introduce a nonparametric dependence of the dynamic POT parameters on our covariates, we follow Chavez-

[7] A likelihood ratio test can reject the Model A in favor of Model B if Model A is nested in Model B. The test is calculated as $LR = -2[\ell(A) - \ell(B)]$. LR is χ^2 distributed with the degrees of freedom being equal to the additional number of parameters in B when compared with A. Additionally, we compare models based on their AIC value, where among several models the one with lowest AIC value is preferred (see e.g. Hambuckers, Kneib et al. (2018) for a discussion on AIC values in generalized additive models).

TABLE 4 U.S. data: Model selection

$u = 50\%$ quantile, $N_u = 3668$ excesses			LRT	AIC	Selection
Frequency parameter, ρ					
(1)	$\rho = Z\overrightarrow{\vartheta}_\rho$			10,203	
(2)	$\rho = Z\overrightarrow{\vartheta}_\rho + \beta_1 x$		(1) : (2) × ($p = .03$)	10,201	
(3)	$\rho = Z\overrightarrow{\vartheta}_\rho + g_\rho^{(1)}(x)$		–	–	
(4)	$\rho = Z\overrightarrow{\vartheta}_\rho + \beta_2 s$		(1) : (4) ✓ ($p < .00$)	10,100	✓
(5)	$\rho = Z\overrightarrow{\vartheta}_\rho + h_\rho^{(5)}(s)$		(4) : (5) × ($p = .01$)	10,095	
Severity parameter, ξ					
(6)	$\xi = Z\overrightarrow{\vartheta}_\xi$	$\nu = Z\overrightarrow{\vartheta}_\nu$		18,008	
(7)	$\xi = Z\overrightarrow{\vartheta}_\xi + \gamma_1 x$	$\nu = Z\overrightarrow{\vartheta}_\nu$	(6) : (7) ✓ ($p < .00$)	9493	✓
(8)	$\xi = Z\overrightarrow{\vartheta}_\xi + g_\xi^{(1)}(x)$	$\nu = Z\overrightarrow{\vartheta}_\nu$	–	–	
(9)	$\xi = Z\overrightarrow{\vartheta}_\xi + \gamma_1 x + \gamma_2 s$	$\nu = Z\overrightarrow{\vartheta}_\nu$	(7) : (9) × ($p > .99$)	9747	
(10)	$\xi = Z\overrightarrow{\vartheta}_\xi + \gamma_1 x + h_\xi^{(1)}(s)$	$\nu = Z\overrightarrow{\vartheta}_\nu$	–	–	
Severity parameter, ν					
(11)	$\xi = Z\overrightarrow{\vartheta}_\xi + \gamma_1 x$	$\nu = Z\overrightarrow{\vartheta}_\nu$		22,710	
(12)	$\xi = Z\overrightarrow{\vartheta}_\xi + \gamma_1 x$	$\nu = Z\overrightarrow{\vartheta}_\nu + \delta_1 x$	(11) : (12) ✓ ($p < .00$)	22,686	
(13)	$\xi = Z\overrightarrow{\vartheta}_\xi + \gamma_1 x$	$\nu = Z\overrightarrow{\vartheta}_\nu + g_\nu^{(1)}(x)$	–	–	
(14)	$\xi = Z\overrightarrow{\vartheta}_\xi + \gamma_1 x$	$\nu = Z\overrightarrow{\vartheta}_\nu + \delta_1 x + \delta_2 s$	(12) : (14) ✓ ($p < .00$)	22,505	✓
(15)	$\xi = Z\overrightarrow{\vartheta}_\xi + \gamma_1 x$	$\nu = Z\overrightarrow{\vartheta}_\nu + h_\nu^{(1)}(s)$	–	–	

Note: The model selection for the frequency parameter ρ and the severity parameters ξ and ν is displayed. For each parameter, we consecutively expand the model by including the company covariate and the time covariate, first parametrically and then nonparametrically. In the fourth column, we report the results of a LRTat a 1% significance level. We denote a significant difference in the likelihood of the two models by "✓" and an insignificant difference by "×." We also provide the corresponding p value in parentheses. Whenever a df-AIC plot indicates that the number of knots of the utilized natural cubic spline is equal to 1 (corresponding to linearity), we do not perform an LRT and indicate this by "–." The fifth column provides the Akaike's information criterion (AIC) value for each model to indicate the model complexity. In the final column, we mark the selected model by "✓."

Abbreviations: AIC, Akaike's information criterion; LRT, likelihood ratio test.

Demoulin et al. (2016, p. 753) and determine the number of knots of the utilized natural cubic spline according to the AIC.

We begin with the specification of the frequency parameter ρ. As shown in Table 4, we consecutively expand the model and compare the specifications based on an LRT. Our baseline model consists of a constant term and the control variables presented in Section 3.1.2. We denote this model by $\rho = Z\overrightarrow{\vartheta}_\rho$ with a matrix Z, consisting of ones and the controls. First, we test the null hypothesis of this baseline model (Model (1)) against the alternative of a parametric inclusion of the covariate number of policies $\rho = Z\overrightarrow{\vartheta}_\rho + \beta_1 x$ (Model (2)) via an LRT. At a 1% significance level, we cannot reject the null hypothesis, meaning that the two likelihoods of Model (1) and Model (2) do not differ by more than sampling error. In the next step, we find that a nonparametric inclusion of the number of policies (Model (3)) does not lead to a marked improvement compared with the parametric inclusion of the number of policies.[8] Subsequently, we include time in a parametric way (Model (4)) and test it against Model (1). The LRT indicates a significant difference in the likelihoods of Model (1) and Model (4). Finally, we vary the degrees of freedom df of the natural cubic spline, $h_\rho^{(df)}(s)$, to

[8] A df-AIC plot (similar to Figure B1 in Supporting Information Appendix B) facilitates $df = 1$ for the function $g_\rho^{(df)}(x)$, which corresponds to the linear model specification of Model (2).

TABLE 5 U.S. data: Threshold selection

Threshold, u	N_u	Model for ξ	Model for ν	Q-Q plot
0% quantile	7336	$\xi = Z\vec{\vartheta}_\xi + g_\xi^{(2)}(x) + h_\xi^{(3)}(s)$	$\nu = Z\vec{\vartheta}_\nu + h_\nu^{(6)}(s)$	×××
10% quantile	6602	$\xi = Z\vec{\vartheta}_\xi + \gamma_1 x$	$\nu = Z\vec{\vartheta}_\nu + \delta_2 s$	××
20% quantile	5869	$\xi = Z\vec{\vartheta}_\xi + \gamma_1 x$	$\nu = Z\vec{\vartheta}_\nu + \delta_2 s$	××
30% quantile	5135	$\xi = Z\vec{\vartheta}_\xi + \gamma_2 s$	$\nu = Z\vec{\vartheta}_\nu + \delta_2 s$	×
40% quantile	4402	$\xi = Z\vec{\vartheta}_\xi + \gamma_1 x$	$\nu = Z\vec{\vartheta}_\nu + \delta_1 x + \delta_2 s$	×
50% quantile	3668	$\xi = Z\vec{\vartheta}_\xi + \gamma_1 x$	$\nu = Z\vec{\vartheta}_\nu + \delta_1 x + \delta_2 s$	✓

Note: The threshold selection procedure is provided. For different selected thresholds u, the number of observations N_u above the chosen threshold and the chosen model specification for the severity parameters ξ and ν are provided. The last column contains the observed goodness of fit in a Q-Q plot based on "×" for bad quality and "✓" for good quality of fit.

obtain the model with the lowest AIC value. Following Chavez-Demoulin et al. (2016, p.753), we plot the AIC value over the degrees of freedom ($df \in 1, 2, ..., 10$), which supports $df = 5$, as an additional increase in the degrees of freedom does not lead to a smaller AIC.[9] We then test Model (4) against this nonparametric inclusion of time, which, however, is not supported by the LRT test at a 1% significance level.

Repeating this procedure for the severity parameters ξ (Models (6)–(10)) and ν (Models (11)–(15)) leads to the following model specification:

$$\rho(x, s) = Z\vec{\vartheta}_\rho + \beta_2 s, \tag{14}$$

$$\xi(x, s) = Z\vec{\vartheta}_\xi + \gamma_1 x, \tag{15}$$

$$\nu(x, s) = Z\vec{\vartheta}_\nu + \delta_1 x + \delta_2 s. \tag{16}$$

Given the model specification of the ex-post selected threshold equal to 50% of all cancellation rates, we now turn to the threshold selection (Table 5). We start the analysis at the 0% quantile and proceed in steps of 10 percentage points. For each threshold, we specify the model according to the steps described in Table 4. Subsequently, we estimate the model and assess the model's goodness of fit in a Q-Q plot, in which we plot the model's residuals against the exponential distribution. The chosen threshold is the smallest threshold so that we observe a good fit in the Q-Q plot (Figure 3a). This analysis leads us to select the 50% quantile as our threshold.

3.1.4 | Results

We estimate the model as specified in Equations (14)–(16). Table 6 provides the coefficients obtained from estimating the generalized additive models.[10] The estimated coefficient for the time variable in the model for ρ demonstrates that the likelihood of excesses decreases over time. This is also shown in Figure 2a, which provides boxplots of the parameter ρ for each year. Table 6 and Figure 2b indicate a further negative effect of the portfolio size on the parameter ξ. In Figure 2b, we also display boxplots

[9] Figure B1 in Supporting Information Appendix B displays the corresponding df-AIC plot.

[10] Figure B2 in Supporting Information Appendix B additionally displays the parameters in dependence of the covariates time and number of policies, fixing all other variables equal to their median values.

TABLE 6 U.S. data: Generalized additive model output

	Dependent variable		
	ρ	ξ	ν
Parametric coefficients	**Coeff. (std. err.)**	**Coeff. (std. err.)**	**Coeff. (std. err.)**
Policy covariate (m)		−0.068***	0.020
		(0.008)	(0.043)
Time	−0.010***		−0.009
	(0.001)		(0.011)
Share of new business	0.561***	−1.019***	4.055***
	(0.041)	(0.057)	(0.554)
Share of term life	0.232***	0.284***	0.445*
	(0.020)	(0.041)	(0.239)
Interest rate changes	0.938	0.024	−2.761
	(1.104)	(2.282)	(13.517)
Intercept	20.286***	0.201***	13.742
	(1.757)	(0.018)	(21.781)
Observations	7336	3668	3668
Log likelihood	−5050	−5186	−11,253

Note: The output of the generalized additive model estimation for the parameters ρ, ξ, and ν, according to the Equations (14)–(16) are displayed. For significance, *, **, and *** indicate the 10%, 5%, and 1% levels, respectively.

for the estimated parameters ξ for each decile of the policy covariate rather than showing the estimated parameters ξ for each value of the policy covariate. In most cases, the shape parameter is positive and implies a support between 0 and positive infinity, whereas cancellation rates are bounded between 0 and 1. Even though our model implies an unbounded support of the GPD, the probability of excesses above 1 is small.[11] Finally, Table 6 provides the estimated coefficient for the parameter ν, and Figure 2c,d display the corresponding boxplots of the estimates for β in the corresponding year and in the decile of the policy covariate, respectively.

For each boxplot of Figure 2, we also provide pointwise two-sided 95% confidence intervals for the displayed median. Whereas we calculate asymptotic confidence intervals based on the standard errors for the parameter ρ, we employ a bootstrap procedure for obtaining confidence intervals for ξ and β. We compute bootstrap confidence intervals by slightly adjusting the approach presented in Chavez-Demoulin et al. (2016), which is based on the post-backend bootstrap of Chavez-Demoulin and Davison (2005). First, we resample the model's residuals with replacements in groups that exhibit the same decile of portfolio size, decile of share of new business, decile of share of term life business, interest rate change, and same year. Second, we calculate new excesses based on the original estimates for ξ and β. Third, we apply the original model with fixed covariates and control variables to obtain new estimates for the severity parameters ξ and β. We repeat this procedure 1,000 times and obtain pointwise two-sided 95% confidence intervals for each parameter estimate by calculating the 2.5% and 97.5% empirical quantiles of the corresponding 1000 estimated values.

[11]Based on the estimates for the shape parameter and the scale parameter, we calculated the probability of cancellation rates exceeding one in our model. Given all these excess probabilities, we find that the empirical 95% quantile of the excess probabilities is smaller than 0.2%.

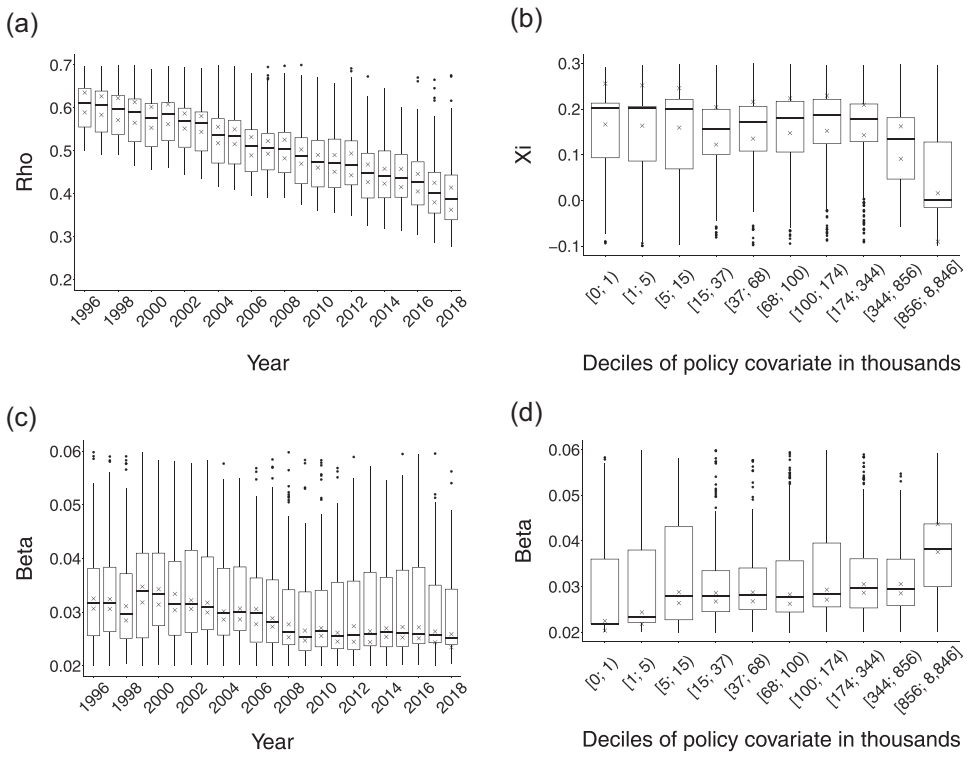

FIGURE 2 U.S. data: Parameter estimates. (a) Boxplot of estimates for ρ over time, (b) boxplot of estimates for ξ for the deciles of the number of policies, (c) boxplot of estimates for β over time, and (d) boxplot of estimates for β for the deciles of the number of policies. Boxplots of the parameter estimates for ρ, ζ, and β in dependence of time and the deciles of the policy covariate are displayed. In addition, "×" denotes pointwise two-sided 95% confidence intervals for the median of the parameter estimates shown in the corresponding boxplot

Given the estimated parameters, we assess the model's goodness of fit and calculate 99.5% quantiles of the cancellation rate observations and corresponding pointwise two-sided 95% confidence intervals (Figure 3). Figure 3a displays a Q-Q plot, in which we plot the selected model's residuals against the exponential distribution. The good fit observed in this plot supports our model selection. Further, Figure 3b provides a boxplot of the estimated 99.5% quantiles for each year between 1996 and 2018. We observe a slight negative trend in the median of 99.5% quantiles over time, indicating that the risk of high cancellation rates decreased over time (consistent with the coefficients of the time parameter in Table 6 and the boxplot in Figure 2a). The boxplots in each year further show substantial heterogeneity in 99.5% quantiles. This suggests that life insurers are exposed heterogeneously to high cancellation rates.

The heterogeneity is further affirmed by examining the 99.5% quantiles in dependence of the number of policies. We provide boxplots of these estimates in Figure 3c for the deciles of the number of policies. In line with Table 6 and Figure 2, Figure 3c indicates that companies with fewer policies in force exhibit higher 99.5% quantiles. In contrast to small life insurance companies, which exhibit 99.5% quantiles as high as 40% (see the upper quartile in the boxplot), the companies in the highest decile of the policy covariate exhibit 99.5% quantiles of around 20%. High cancellation rates are additionally dependent on the share of new business in dependence of existing business and the product portfolio (Figure 3d,e). We find that a high share of new business activity results in higher cancellation rates, reaching levels of approximately 60% in the highest decile of this variable. In

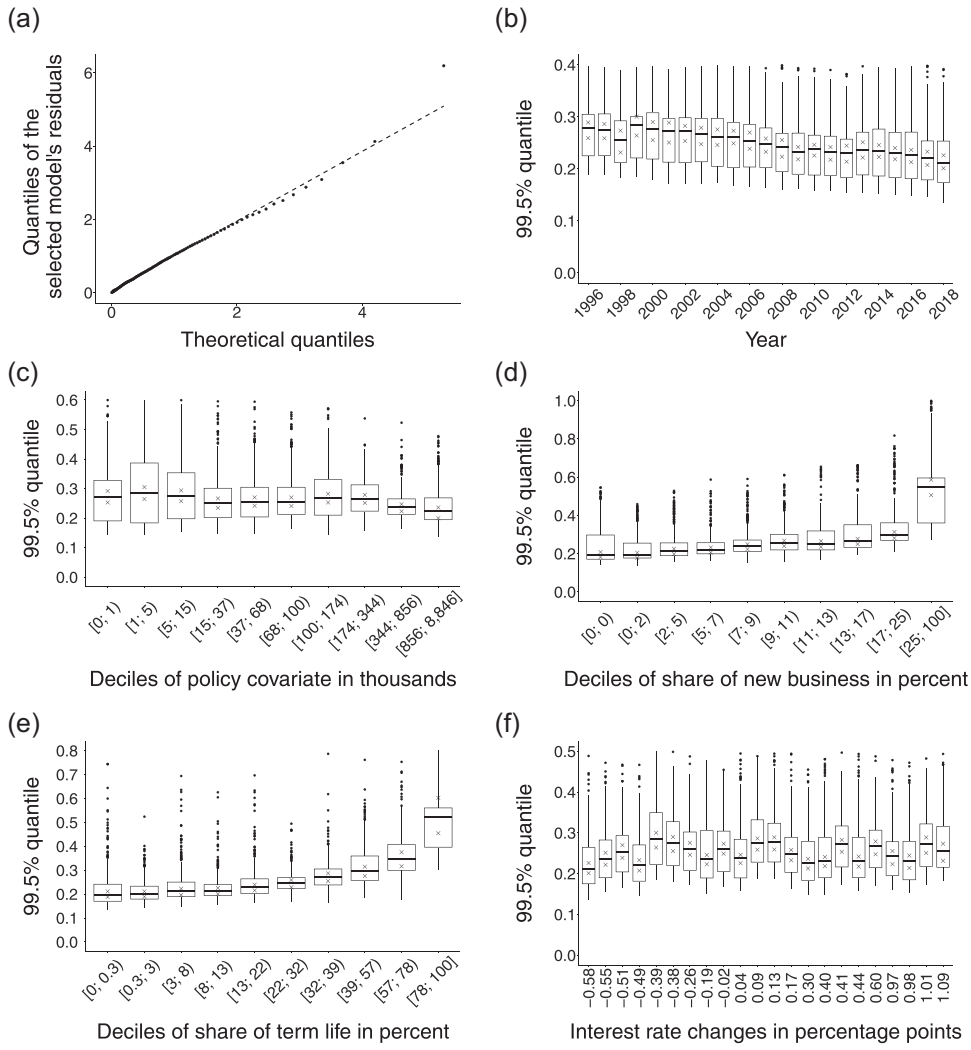

FIGURE 3 U.S. data: Estimation results. (a) Q-Q plot: Goodness of fit; (b) boxplot of 99.5% quantiles over time; (c) boxplot of 99.5% quantiles for the deciles of the number of policies; (d) boxplot of 99.5% quantiles for the deciles of the share of new business; (e) boxplot of 99.5% quantiles for the deciles of the share of term life; (f) boxplot of 99.5% quantiles for the interest rate changes. A goodness of fit plot in (a) and estimation results in (b)–(f) are displayed. The goodness of fit plot is a Q-Q plot based on the selected model's residuals. (b) provides a boxplot of estimated 99.5% quantiles for each year between 1996 and 2018. (c) shows the boxplot of estimated 99.5% quantiles for the deciles of the policy covariate; (d), for the deciles of the share of new business; (e), for the deciles of the share of term life; and (f), for the interest rate changes. The estimation results also include pointwise two-sided 95% confidence intervals for the median of the 99.5% quantiles, indicated by "×" in the corresponding boxplot

Figure 3e, we additionally observe that a portfolio with a high share of term life business is more exposed to high cancellation rates compared with a portfolio that consists mainly of permanent insurance. Although life insurance companies with predominantly permanent life insurance policies exhibit 99.5% quantiles of 20%, we find that life insurers, which focus mainly on term life policies, have 99.5% quantiles up to 50%.

TABLE 7 German data: Summary statistics

Variable	Description	*N*	Mean	Pctl(25)	Median	Pctl(75)
gep^{DE} (EUR)	Gross earned premium	2344	720 m	63 m	216 m	779 m
$inFore^{DE}$ (EUR)	Amount of insurance in force	2344	23 bn	2 bn	9 bn	27 bn
$issued^{DE}$ (EUR)	Amount of insurance issued	2344	2 bn	92 m	525 m	2 bn
$shareIssued^{DE}$	Share of issued policies	2344	11.34%	5.02%	8.59%	13.60%
$interest^{DE}$	Interest rate changes	23	−0.25pp	−0.54pp	−0.33pp	0.07pp
$cancel^{DE}$	Cancellation rates	2344	5.15%	3.35%	4.67%	6.18%

Note: Summary statistics of the variables used in the German analysis of extreme cancellation rates are provided. The data include 122 life insurers between 1996 and 2018.

The previous literature models extreme cancellation rates as a result of changes in interest rates (Barsotti et al., 2016; Loisel & Milhaud, 2011). Figure 3f displays boxplots of the 99.5% quantiles dependent on the interest rate changes in U.S. government bonds with a duration of 10 years. At least in our observed range of interest rate changes, we do not identify any pattern between changes in interest rates and high cancellation rates. In addition to Figure 3f, Table 6 shows that the coefficients of the interest rate variable are insignificant.

3.2 | German life insurance market

3.2.1 | Motivation

Cancellation risk is the second most important risk factor for European life insurers and is, therefore, explicitly addressed in European insurance regulation (EIOPA, 2011). Within its standard formula, Solvency II requires life insurers to calculate additional capital demands for adverse cancellation events, such as a permanent increase or decrease of future cancellation rates. In addition, life insurers must apply a one-time shock mass lapse scenario with cancellation rates equal to 40%. Interestingly, this level of 40% has not been predicted by utilizing data but is based only on expert judgment. In contrast to the expert judgment-based approach in Solvency II's standard model, our approach can use data to calibrate an empirically justified mass cancellation scenario. For this purpose, we again use available panel data at the company level.

3.2.2 | Company-level, panel data

The German Federal Financial Supervisory Authority (BaFin) collects information on all German life insurers subject to the Solvency II regulation. German life insurers predominantly offer endowment policies, term life, occupational disability, annuity, unit-linked insurance, and group life insurance.[12]

[12]An endowment policy either pays the sum insured in the case of death or if the insured person is still at life at the end of the contract period. A term life policy pays the sum insured if the insured person dies within the contract period. An occupational disability policy pays an annuity until retirement if the insured person is no longer able to work. An annuity policy pays an annuity until the insured person's death starting at a predefined age. Special forms of annuities are Riester and R'rup policies (see Börsch-Supan & Wilke, 2004 for more details). Finally, a unit-linked policy pays the sum insured if the insured person is still alive at the end of the contract period. The accrued interest is dependent on the performance of stocks or funds.

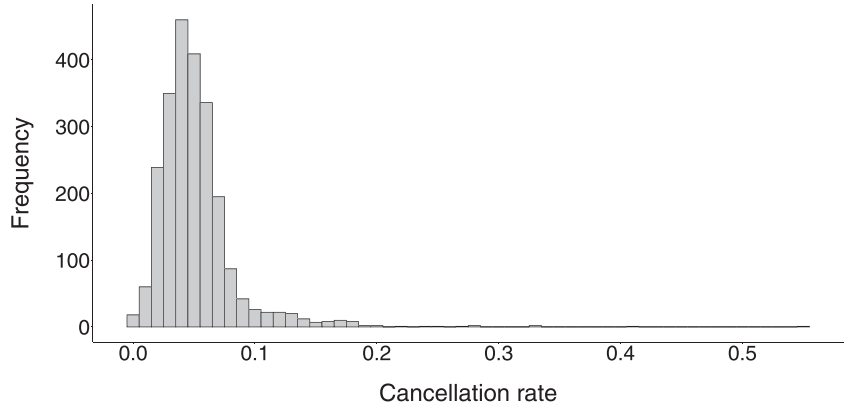

FIGURE 4 German data: Histogram of cancellation rates. The distribution of the 2344 cancellation rates between 1996 and 2018 is displayed

Although we are unable to distinguish between these product categories, Eling and Kiesenbauer (2013) use policyholder-level data and find that, in contrast to other markets, the German life insurance market exhibits only small differences in cancellation rates across product categories. Our data provide gross earned premiums, detailed portfolio information (e.g., amount of insurance in force, amount of insurance issued), and cancellation rates of 122 life insurers between 1996 and 2018. In contrast to the U.S. analysis, we do not apply any filters because the BaFin database includes only active primary insurers, and policies can be revived for only as long as one month. Table 7 provides the summary statistics of our company-level panel data, and Figure 4 provides a histogram of the 2344 cancellations rates. Although the median cancellation rates in both the German and the U.S. data is around 4.7%, U.S. cancellation rates exhibit higher variability. The interquartile range of cancellation rates is 4.34% for the United States and 2.83% for Germany. Even though German cancellation rates are less variable as compared with those of the U.S. market, Figure 4 also shows cancellation rates above 20% in the German market. Because we are again interested in a reliable estimate of the 99.5% quantile of the cancellation rates, employing the dynamic POT method seems particularly suitable, given the few observations in the tail. Due to data availability, we will use the gross earned premium instead of the number of policies as a covariate in the German estimation. We additionally include time as a covariate and the following control variables: (1) We monitor new business activity by the share of issued policies in terms of existing policies ($shareIssued^{DE}$). (2) We control for changes in market interest rates by the first-difference of German government bond yields with a duration of 10 years ($interest^{DE}$).

3.2.3 | Model selection

In the German analysis, we select a threshold equal to the 65% quantile of all cancellation rates, corresponding to $u = 5.48\%$ and 821 excesses. Again, the specification of the threshold is done via backward induction. For each threshold equal to the deciles of all cancellation rates, we perform model selection for the severity of extreme cancellation rates. We then choose the lowest threshold with a good model fit, according to a Q-Q plot. As in the U.S. analysis, we specify the model for the frequency parameter ρ and severity parameters ξ and ν based on LRTs and the AIC value (Table 8).

TABLE 8 German data: Model selection

$u = 65\%$ quantile, $N_u = 821$ excesses			LRT	AIC	Selection
Frequency parameter, ρ					
(1)	$\rho = Z\vec{\vartheta}_\rho$			3139	
(2)	$\rho = Z\vec{\vartheta}_\rho + \beta_1 x$		(1) : (2) ✓ ($p < .01$)	3131	
(3)	$\rho = Z\vec{\vartheta}_\rho + g_\rho^{(1)}(x)$		–	–	
(4)	$\rho = Z\vec{\vartheta}_\rho + \beta_1 x + \beta_2 s$		(2) : (4) ✓ ($p < .00$)	2969	
(5)	$\rho = Z\vec{\vartheta}_\rho + \beta_1 x + h_\rho^{(2)}(s)$		(4) : (5) ✓ ($p < .00$)	2959	✓
Severity parameter, ξ					
(6)	$\xi = Z\vec{\vartheta}_\xi$	$\nu = Z\vec{\vartheta}_\nu$		2201	
(7)	$\xi = Z\vec{\vartheta}_\xi + \gamma_1 x$	$\nu = Z\vec{\vartheta}_\nu$	(6) : (7) × (p > .99)	2875	
(8)	$\xi = Z\vec{\vartheta}_\xi + g_\xi^{(2)}(x)$	$\nu = Z\vec{\vartheta}_\nu$	(6) : (8) ✓ ($p < .00$)	2136	
(9)	$\xi = Z\vec{\vartheta}_\xi + g_\xi^{(2)}(x) + \gamma_2 s$	$\nu = Z\vec{\vartheta}_\nu$	(8) : (9) ✓ ($p < .00$)	2077	✓
(10)	$\xi = Z\vec{\vartheta}_\xi + g_\xi^{(2)}(x) + h_\xi^{(1)}(s)$	$\nu = Z\vec{\vartheta}_\nu$	–	–	
Severity parameter, ν					
(11)	$\xi = Z\vec{\vartheta}_\xi + g_\xi^{(2)}(x) + \gamma_2 s$	$\nu = Z\vec{\vartheta}_\nu$		7563	
(12)	$\xi = Z\vec{\vartheta}_\xi + g_\xi^{(2)}(x) + \gamma_2 s$	$\nu = Z\vec{\vartheta}_\nu + \delta_1 x$	(11) : (12) ✓ ($p < .00$)	7342	✓
(13)	$\xi = Z\vec{\vartheta}_\xi + g_\xi^{(2)}(x) + \gamma_2 s$	$\nu = Z\vec{\vartheta}_\nu + g_\nu^{(1)}(x)$	–	–	
(14)	$\xi = Z\vec{\vartheta}_\xi + g_\xi^{(2)}(x) + \gamma_2 s$	$\nu = Z\vec{\vartheta}_\nu + \delta_1 x + \delta_2 s$	(12) : (14) × ($p > .99$)	7427	
(15)	$\xi = Z\vec{\vartheta}_\xi + g_\xi^{(2)}(x) + \gamma_2 s$	$\nu = Z\vec{\vartheta}_\nu + \delta_1 x + h_\nu^{(1)}(s)$	–	–	

Note: The model selection for the frequency parameter ρ and the severity parameters ξ and ν are presented. For each parameter, we consecutively expand the model by including the company covariate and the time covariate, first parametrically and then nonparametrically. In the fourth column, we report the result of an LRT at a 1% significance level. We denote a significant difference in the likelihood of two models by "✓" and an insignificant difference by "×." We also provide the corresponding p value in parentheses. Whenever a df-AIC plot supports that the number of knots of the utilized natural cubic spline is equal to 1 (corresponding to linearity), we do not perform an LRT and indicate this by "–." The fifth column provides the AIC value for each model to indicate the model complexity. In the final column we mark the selected model by "✓."

Abbreviations: AIC, Akaike's information criterion; LRT, likelihood ratio test.

We begin with the specification of the frequency parameter ρ, and our baseline is again a model, $\rho = Z\vec{\vartheta}_\rho$, with a matrix Z that consists of ones and the control variables of Section 3.2.2. We test the null hypothesis of this baseline model (Model (1)) against the alternative of a parametric inclusion of the covariate gross earned premium, $\rho = Z\vec{\vartheta}_\rho + \beta_1 x$ (Model (2)) via an LRT. At a 1% significance level, we can reject the null hypothesis of a constant model. Although we cannot reject the null hypothesis of Model (2) when comparing it to a nonparametric inclusion of the gross earned premium (Model (3)), a parametric inclusion of time is supported (Model (4)). Finally, a df-AIC plot and an LRT test facilitate to specify the parameter ρ as a natural cubic spline function of time with two degrees of freedom (Model (5)). We repeat this procedure for the severity parameters ξ (Models (6)–(10)) and ν (Models (11)–(15)) and finally obtain the following model specification:[13]

$$\rho(x, s) = Z\vec{\vartheta}_\rho + \beta_1 x + h_\rho^{(2)}(s), \tag{17}$$

[13]The dependence of the parameter ξ on time is unusual. In most applications, the shape parameter does not depend on time (see e.g., Chavez-Demoulin et al., 2016). However, in our empirical setting, this does not lead to a switch between finite/infinite first moment of the GPD over time.

TABLE 9 German data: Threshold selection

Threshold, u	N_u	Model for ξ	Model for ν	Q-Q plot
10% quantile	2110	$\xi = Z\vec{\vartheta}_\xi + g_\xi^{(4)}(x)$	$\nu = Z\vec{\vartheta}_\nu$	×××
20% quantile	1881	$\xi = Z\vec{\vartheta}_\xi$	$\nu = Z\vec{\vartheta}_\nu + h_\nu^{(2)}(s)$	×××
30% quantile	1641	$\xi = Z\vec{\vartheta}_\xi$	$\nu = Z\vec{\vartheta}_\nu + \delta_1 x + \delta_2 s$	××
40% quantile	1407	$\xi = Z\vec{\vartheta}_\xi + \gamma_2 s$	$\nu = Z\vec{\vartheta}_\nu + g_\nu^{(3)}(x) + \delta_2 s$	××
50% quantile	1172	$\xi = Z\vec{\vartheta}_\xi + g_\xi^{(3)}(x) + \gamma_2 s$	$\nu = Z\vec{\vartheta}_\nu + h_\nu^{(2)}(s)$	×
55% quantile	1055	$\xi = Z\vec{\vartheta}_\xi + g_\xi^{(3)}(x)$	$\nu = Z\vec{\vartheta}_\nu + h_\nu^{(2)}(s)$	×
60% quantile	938	$\xi = Z\vec{\vartheta}_\xi + g_\xi^{(2)}(x)$	$\nu = Z\vec{\vartheta}_\nu + \delta_2 s$	×
65% quantile	821	$\xi = Z\vec{\vartheta}_\xi + g_\xi^{(2)}(x) + \gamma_2 s$	$\nu = Z\vec{\vartheta}_\nu + \delta_1 x$	✓

Note: The threshold selection procedure is presented. For different selected thresholds u, the number of observations N_u above the chosen threshold and the chosen model specification for the severity parameters ξ and ν are presented. The last column indicates the observed goodness of fit in a Q-Q plot based on "×" for bad quality and "✓" for good quality of fit.

$$\xi(x, s) = Z\vec{\vartheta}_\xi + g_\xi^{(2)}(x) + \gamma_2 s, \tag{18}$$

$$\nu(x, s) = Z\vec{\vartheta}_\nu + \delta_1 x. \tag{19}$$

We now return to the threshold selection (Table 9). The chosen threshold is the smallest threshold, so we observe a good fit in the Q-Q plot (Figure 6a). This analysis supports our selection of the 65% quantile as our threshold.[14]

3.2.4 | Results

We estimate the model as specified in Equations (17)–(19). Table 10 provides the parametric coefficients and the degrees of freedom of the smooth terms obtained from estimating the generalized additive models.[15]

The estimated coefficient for the policy covariate in the model for ρ demonstrates that the likelihood of excesses decreases with increasing portfolio size. This is also shown in Figure 5b, which provides boxplots of the parameter ρ for each decile of the gross earned premium. Figure 5a also displays a negative relationship between the likelihood of excesses over time. Although the parameter ρ decreases over time, Table 10 and Figure 5c indicate a positive relationship between ξ and time. Consistent with the U.S. analysis, Table 10 and Figure 5d shows a further negative effect of the portfolio size on the parameter ξ. Again, even though positive values of ξ imply an unbounded support of the GPD, the probability of excesses above 1 is small.[16] Finally, Table 10 provides the estimated coefficient for the parameter ν, and Figure 5e,f display the corresponding

[14] We start with the 10% quantile as our threshold and proceed in steps of 10 percentage points. After the 50% quantile, we start moving in 5 percentage points increments to still have a reasonably large sample size.

[15] Figure B3 in Supporting Information Appendix B additionally displays the parameters in dependence of the covariates time and gross earned premium, fixing all other variables equal to their median values.

[16] Based on the estimates for the shape parameter and the scale parameter, we calculated the probability that cancellation rates exceed one in our model. Given all these excess probabilities, we find that the empirical 95% quantile of the excess probabilities is smaller than 0.2%.

TABLE 10 German data: Generalized additive model output

	Dependent variable		
	ρ	ξ	ν
Parametric coefficients	**Coeff. (std. err.)**	**Coeff. (std. err.)**	**Coeff. (std. err.)**
Policy covariate (in bn)	−0.010*		−0.397***
	(0.006)		(0.077)
Time		0.026***	
		(0.003)	
Share of new business	0.290***	−0.208	2.463*
	(0.078)	(0.155)	(1.442)
Interest rate changes	−1.282	1.420	−12.083
	(1.629)	(2.472)	(23.818)
Intercept	0.322***	−51.932***	−4.022***
	(0.015)	(6.878)	(0.311)
Smooth terms	***df***	***df***	***df***
Policy covariate (in bn)		2***	
Time	2***		
Observations	2344	821	821
Log likelihood	−1479	−860	−2670

Note: The output of the generalized additive model estimation for the parameters ρ, ξ, and ν, according to Equations (17)–(19) are presented. For significance, *, **, and *** indicate the 10%, 5%, and 1% levels, respectively.

boxplots of the estimates of β in the corresponding year and in the decile of the policy covariate, respectively.

Given the estimated parameters, we again assess the model's goodness of fit and calculate 99.5% quantiles of the cancellation rate observations and corresponding pointwise two-sided 95% confidence intervals (Figure 6). Figure 6a shows the model's goodness of fit in a Q-Q plot and supports our model selection. Figure 6b presents boxplots for the estimated 99.5% quantiles for each year between 1996 and 2018. The median of 99.5% quantiles never exceeds 22% (the corresponding upper bound of the confidence interval never exceeds 25%) and exhibits a decreasing trend over this time period. The higher level of extreme cancellation rates at the beginning of the observation period is likely a result of the dot-com bubble. During this period, some German life insurers faced solvency issues and, thus, a decline in trust in their company (e.g., "Mannheimer Leben," Protektor, 2014). In comparison with the U.S. analysis, however, we observe less heterogeneity in 99.5% quantiles for each year.

Further, Figure 6c displays boxplots of the 99.5% quantiles dependent on the deciles of the gross earned premium. Figure 6c indicates that the level of extreme cancellation rates is decreasing in the gross earned premium (in line with Table 10 and Figure 5). Although the companies in the lowest decile of the policy covariate exhibit 99.5% quantiles of about 20%–25%, this risk lies at around 10% for life insurers in the largest decile of the policy

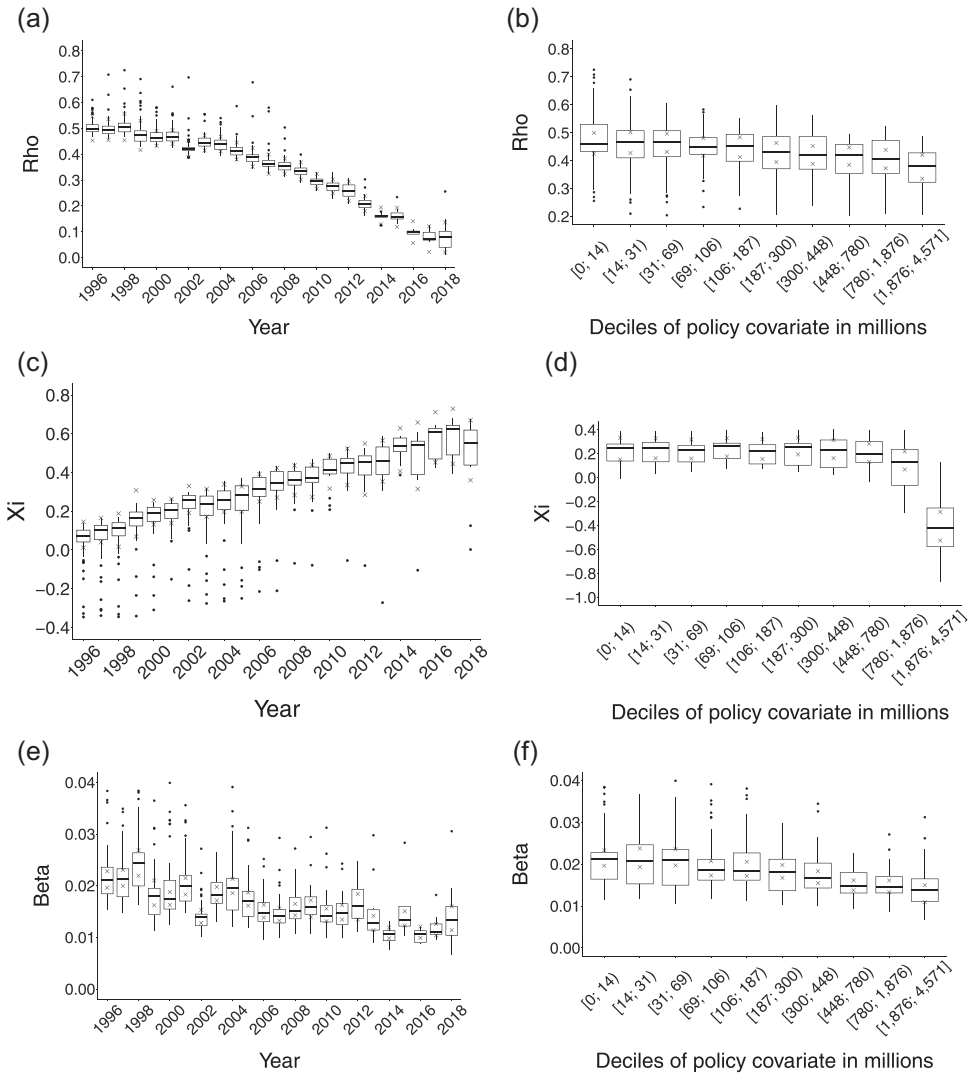

FIGURE 5 German data: Parameter estimates. (a) Boxplot of estimates for ρ over time; (b) boxplot of estimates for ρ for the deciles of the gross earned premium; (c) boxplot of estimates for ξ over time; (d) boxplot of estimates for ξ for the deciles of the gross earned premium; (e) boxplot of estimates for β over time; (f) boxplot of estimates for β for the deciles of the gross earned premium. Boxplots of the parameter estimates for ρ, ξ, and β in dependence of time and the deciles of the policy covariate are displayed. In addition, "×" denotes pointwise two-sided 95% confidence intervals for the median of the parameter estimates shown in the corresponding boxplot

covariate. As in the U.S. analysis, Figure 6d demonstrates that a high share of new business activity results in higher cancellation rates, which are estimated to reach as high as 35%. In addition, in Figure 6e, we observe, in line with the U.S. analysis, no pattern between changes in interest rates and high cancellation rates. This is also consistent with the insignificant coefficients of the interest rate variable in Table 10.

Our dynamic POT estimation also enables us to calculate the confidence level that corresponds to a cancellation rate of 40%. In each year between 1996 and 2018, the median

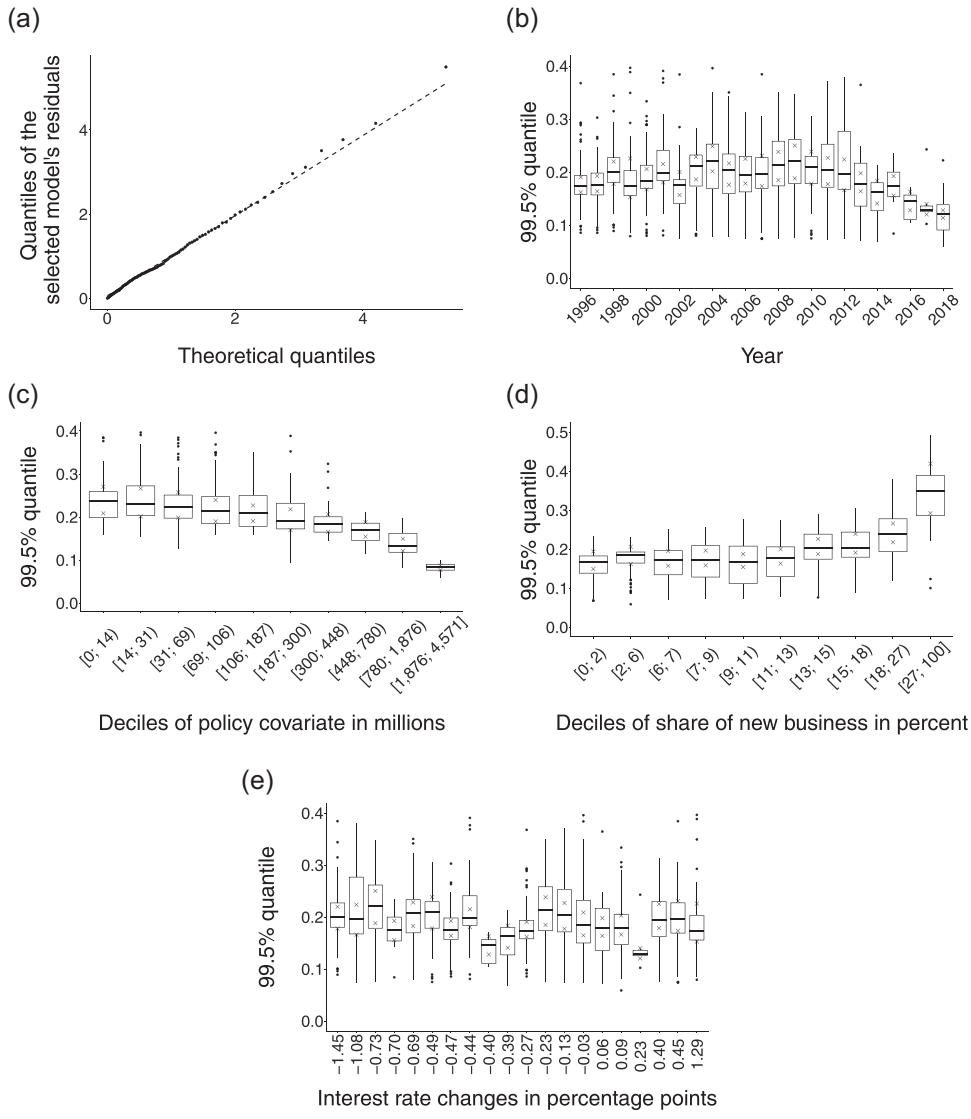

FIGURE 6 German data: Estimation results. (a) Q-Q plot: Goodness of fit; (b) boxplot of 99.5% quantiles over time; (c) boxplot of 99.5% quantiles for the deciles of the gross earned premium; (d) boxplot of 99.5% quantiles for the deciles of the share of new business; (e) boxplot of 99.5% quantiles for the interest rate changes. A goodness of fit plot in (a) and estimation results in (b)–(e) are displayed. The goodness of fit plot is a Q-Q plot based on the selected model's residuals. (b) provides a boxplot of estimated 99.5% quantiles for each year between 1996 and 2018. (c) shows the boxplot of estimated 99.5% quantiles for the deciles of the policy covariate; (d), for the deciles of the share of new business; and (e), for the interest rate changes. The estimation results also include pointwise two-sided 95% confidence intervals for the median of the 99.5% quantiles indicated by "×" in the corresponding boxplot

of the implied confidence levels lies above 99.9%. We thus conclude that Solvency II's mass lapse assumption corresponds to the 99.9% quantile rather than the 99.5% quantile, on which the solvency capital requirement in European insurance regulation is usually based.

4 | DISCUSSION

4.1 | Implications for the modeling of extreme cancellation rates

Our results allow several conclusions. Although the U.S. and the German life insurance market are different in terms of their offered products, we find that the effects of the covariates (portfolio size, time) and control variables (share of new business, interest rate changes) are consistent across both markets. This demonstrates the validity of the employed estimation method. In both the U.S. and the German life insurance market, we find that the risk of extreme cancellation rates is decreasing in proportion to the size of the company's portfolio. We also observe a positive relationship between new business activity and extreme cancellation rates. Because policies in early policy years are subject to higher cancellation rates, a high share of new business raises the likelihood and severity of an extreme cancellation event.

We also focus on the effect of changes in interest rates on extreme cancellation rates. The previous literature models extreme cancellation rates as a result of increasing interest rates (Barsotti et al., 2016; Loisel & Milhaud, 2011). Our analysis shows no pattern between changes in interest rates and high cancellation rates. However, interest rates are mainly decreasing in our time period, which does not allow us to test the effect of increases in interest rates on extreme cancellation rates. The U.S. analysis further indicates the importance of the product portfolio on cancellation rates, at least in certain markets. Although life insurance companies with predominantly permanent life insurance policies exhibit 99.5% quantiles of 20%, we find that life insurers, which focus mainly on term life policies, have 99.5% quantiles up to 50%. Renshaw and Haberman (1986), Cerchiara et al. (2009), and Milhaud et al. (2011) find that this also holds at normal cancellation rate levels, using data from Scotland, Italy, and Spain, respectively. With the German data, we are, unfortunately, unable to distinguish between product categories. Eling and Kiesenbauer (2013), however, show that, in contrast to other markets, the German life insurance market exhibits only small differences in cancellation rates across product categories. Still, a thorough analysis of the German market by product type is a promising area for future research.

Interestingly, the results above identify dependencies of a mass cancellation event on company characteristics, which, so far, have not been taken into account in Solvency II. The adverse cancellation rate scenarios used to calculate the solvency capital requirement do not distinguish between small and large insurance companies, run-off companies without new business, or companies that exhibit a high share of new business. These scenarios also are not dependent on the product type. We further find that the current scenario of 40% for a mass cancellation event in Solvency II's standard model has, at least for the German life insurance market, no empirical foundation. According to our analysis, a cancellation rate of 40% corresponds to the 99.9% quantile. The 99.5% quantile, on which the solvency capital requirement in European insurance regulation is usually based, lies at approximately 20%–25% (see the median of 99.5% quantiles and the corresponding upper bound of the confidence interval in Figure 6b).

4.2 | Implications for the solvency capital requirement calculation

The solvency capital requirement for European life insurers is sensitive to changes in cancellation rates and is particularly sensitive to a mass cancellation scenario (EIOPA, 2011). South African insurer Old Mutual (2016) reports that the solvency capital requirement for its business in Europe for a mass cancellation event is equal to 500 million pounds. Austrian life

insurer UNIQA (2017) estimates the capital demand for this risk to be 262 million Euros. European insurance regulation define the required capital for an adverse scenario, such as a mass cancellation event, as the reduction in expected future profits due to deviations from the company's best-estimate assumptions. In the following, we illustrate the sensitivity of a term life policy's expected profitability to a mass cancellation event.[17]

The expected profitability, called technical provision (TP), at time τ is defined as the difference between the expected present value of future cash outflows at time τ and the expected present value of future cash inflows at time τ. To calculate the technical provision under Solvency II, we discount expected future cash flows by discount factors, which can be calculated from the spot rates provided by EIOPA (2019). For example, a negative technical provision at contraction ($\tau = 0$) means that the contract generates overall expected future profits for the insurer. In our illustration, we consider a 40-year-old insured who purchases a term life policy with a contract period of 20 years and a sum insured of 150,000 Euros (death benefit).[18]

The insurer expects cash inflows in terms of a level premium at time $\tau = 0, 1, ..., 19$ and is obliged to pay the specified death benefit at time $\tau = 1, 2, ..., 20$ if the insured dies. Based on the above assumptions and using mortality rates and a guaranteed interest rate of 0.9%, according to the German Actuarial Association, the policy's premium is equal to 495.37 Euros. During the contract period, the policyholder can exercise the option to cancel the policy, which terminates future premium payments as well as the death benefit. In the calculation of expected future profits, an insurer assumes best-estimate cancellation rates for each year within the contract period. We use best-estimate cancellation rates, which are typical for a term life policy with a contract period of 20 years in the German market (Milbrodt & Helbig, 2008). Cancellation rates are high in the first policy years and are decreasing afterward.[19] Figure 7 shows the expected profitability of the policy under best-estimate cancellation assumptions as well as two mass cancellation scenarios: We (1) consider Solvency II's 40% mass lapse shock and (2) calculate the expected profitability with a 25% mass lapse shock, as it is supported by our analysis in Section 3.2.4.

According to European insurance regulation, we calculate the expected profitability at time τ under a mass cancellation event by taking the best-estimate cancellation rates but replacing the cancellation rate at time $\tau + 1$ by the mass lapse shock (40% according to Solvency II, and 25% according to our calculation). The figure shows that, for the first eight contract years, the technical provision (see the solid line) is negative, meaning that the policy generates an expected future profit for the insurer. Afterward, the technical provision is positive, indicating that the policy generates an expected future loss. It is thus most profitable for the insurance company if the insured does not terminate her term life policy during the first contract years but, rather, cancels it as soon as the technical provision becomes positive in later contract years. Following this logic, it is reasonable that a mass lapse scenario (see the dashed lines) leads to a higher technical provision during the first six years and to a lower technical provision after this period when compared with the best-estimate scenario.

Given the technical provision over time, we can do a back-of-the-envelope calculation of the effect of a mass lapse event on the capital demand for a life insurance portfolio. First, we build a portfolio of term life policies by assuming a constant new business volume of 6500 policies over

[17] Supporting Information Appendix A provides the corresponding actuarial calculations in detail.

[18] On average, Germans purchase a home at the age of 40. They usually purchase a term life policy to hedge the risk of an early death, which leaves family members with the debt incurred by building loans. In 2018, the average period for building loans was 20 years and the average sum insured for term policies was 150,000 Euros (BaFin, 2019; SOEP, 2019).

[19] Column 2 of Table A1 in Supporting Information Appendix A displays these best-estimate cancellation rates.

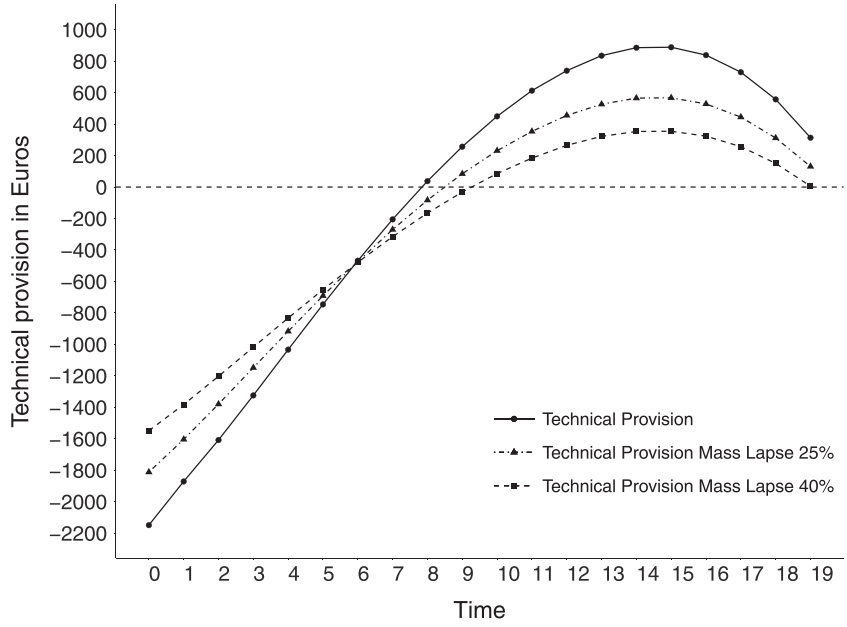

FIGURE 7 Sensitivity of the technical provision to Solvency II's mass lapse scenario. The expected profitability of a term life policy under best-estimate cancellation assumptions and two mass cancellation scenarios is displayed. We (1) consider Solvency II's 40% mass lapse shock and (2) calculate the expected profitability with a 25% mass lapse shock, as it is supported by our analysis in Section 3.2.4

the next 20 years. This figure corresponds to the average number of issued term life policies in Germany in 2018 (BaFin, 2019). At time $\tau = 0$, we have 6500 policies. Considering the above mortality and cancellation rates, we expect that, of these 6500 policies, 6119 policies are left in the portfolio at time $\tau = 1$; then, 5690 policies, at time $\tau = 2$; and, finally, 3204 policies, at time $\tau = 19$. Second, we calculate the expected profitability of the term life policy under best-estimate cancellation rates and under a mass lapse scenario for each point in time. For example, the expected profitability at time $\tau = 0$ is 2149 Euros ($TP(0) = -2,149$ in Figure 7). Assuming a mass cancellation scenario with a shock level of 40% reduces the profitability of this policy to 1549 Euros and, with a shock level of 25%, to 1812 Euros. Performing this calculation for each point in time $\tau = 0, 1, ..., 19$ provides the capital demand for the above portfolio that consists of term life policies. The capital demand is equal to 12,310,895 Euros with a mass lapse shock level of 40% and 6,920,217 Euros with a mass lapse shock level of 25%. Therefore, reducing the shock level for the mass lapse scenario from 40% to 25% can reduce the capital demand for lapse risk by 5 million Euros for this portfolio.[20] This illustration and the figures reported by Old Mutual (2016) and UNIQA (2017) demonstrate that the assumption of a 40% mass lapse shock in the cancellation risk model results in a high capital demand. We thus conclude that the overstatement of the mass lapse event in Solvency II leads to significant over-reserving compared with empirically more justifiable scenarios.

[20]Solvency II defines the capital demand for a portfolio as maximum capital demand required under the following three scenarios: (1) mass lapse shock of 40%, (2) 50% increase in best-estimate cancellation rates, and (3) 50% decrease in best-estimate cancellation rates. In our example, the 25% mass lapse shock still requires more capital than a 50% increase in best-estimate cancellation rates and approximately the same capital as a 50% decrease in best-estimate cancellation rates. Supporting Information Appendix A provides all details of the calculation.

5 | CONCLUSION

We contribute to the literature by assessing the risk of a mass cancellation scenario in life insurance. As extreme cancellation rates are rare events, data sourced from only one insurer do not provide enough observations to assess this tail risk. We thus use the dynamic peaks over the threshold method developed by Chavez-Demoulin et al. (2016) to take quantitative covariates into account. We apply this approach to U.S. data and show that, depending on product type, cancellation rates up to 50% are a good assumption for a mass cancellation scenario in this market. Further, we provide implications for European insurance regulation. We discuss the appropriateness of Solvency II's mass lapse scenario by using German data. Cancellation rates in the range of 20%–25% reflect the risk of a mass cancellation scenario in the German market. This calls into doubt whether the current scenario of a 40% cancellation rate in Solvency II's standard model is empirically justified. In both the United States and Germany, the mass cancellation scenario is dependent on the portfolio size, which provides some validity to the use of (partial) internal models when assessing reserves for mass cancellation scenarios. Finally, the marked differences in the estimated mass cancellation scenarios between the U.S. companies that sell predominantly term life insurance and those that sell predominantly permanent life insurance indicate that the product type has an effect on the appropriate mass cancellation scenario, at least in certain markets. Because national life insurance markets in Europe greatly differ with regard to their dominant products (Standard & Poors, 2018), it can thus be expected that they differ with respect to their appropriate mass cancellation scenario as well.

Our work provides some direction for further research. Even though Eling and Kiesenbauer (2013) report only minor differences in the cancellation rates of different product types in the German market, future research should nevertheless employ the dynamic POT method on a richer data set to analyze the mass cancellation scenario by product type in the German market. Given the differences by product type in the U.S. analysis, different mass cancellation scenarios for different product types could be adequate in the German market as well. Because such data are not publicly available in a panel structure of the entire market, cooperation with a regulating entity, such as BaFin or EIOPA, is likely required. In addition, the lack of available public data for time periods with increasing interest rates allows us to draw conclusions only about economies with falling or stagnant interest rates. Further research on other time periods could provide an estimate for the influence of interest rate development on mass cancellation risk. Finally, the approach of modeling the mass cancellation scenario equally for all European life insurance markets is called into question in our results. An empirical study that uses data from several different national insurance markets in Europe would be able to test it directly.

ACKNOWLEDGMENTS

We have received helpful suggestions from the Editor, one Senior Editor, and two anonymous referees. In addition, we are thankful for thoughtful comments from participants at the 2017 annual meeting of the German Insurance Science Association, the 2017 annual meeting of the American Risk and Insurance Association, the 2018 International Congress of Actuaries, and seminar participants at Temple University. We are indebted to Alexander Braun, Evan Eastman, Pierre Joos, Alexander Kling, Lu Li, Thorsten Moenig, Joëlle Näger, Andreas Richter, Jochen Ruß, Günter Schwarz, and Richard Vierthauer for valuable comments. Open access funding enabled and organized by Projekt DEAL.

ORCID

Francesca Biagini http://orcid.org/0000-0001-9801-5259
Tobias Huber https://orcid.org/0000-0002-9894-9710
Johannes G. Jaspersen https://orcid.org/0000-0002-3599-8988

REFERENCES

ACLI. (2018). *Life insurance fact book*. American Council of Life Insurers. https://www.acli.com/posting/rp18-007

Albizzati, M.-O., & Geman, H. (1994). Interest rate risk management and valuation of the surrender option in life insurance policies. *Journal of Risk and Insurance*, *61*(4), 616-637.

Bacinello, A. (2003). Fair valuation of a guaranteed life insurance participating contract embedding a surrender option. *Journal of Risk and Insurance*, *70*(3), 461-487.

BaFin. (2019). Statistik der Erstversicherungsunternehmen—Lebensversicherung. Bundesanstalt für Finanzdienstleistungsaufsicht. https://www.bafin.de/DE/PublikationenDaten/Statistiken/Erstversicherung/erstversicherung_artikel.html

Balkema, A., & de Haan, L. (1974). Residual life time at great age. *Annals of Probability*, *2*(5), 792–804.

Barsotti, F., Milhaud, X., & Salhi, Y. (2016). Lapse risk in life insurance: Correlation and contagion effects among policyholders' behaviors. *Insurance: Mathematics and Economics*, *71*, 317–331.

Börsch-Supan, A., & Wilke, C. (2004). *The German public pension system: How it was, how it will be* (NBER Working Paper No. 10525).

CEIOPS. (2009). *CEIOPS' advice for level 2 implementing measures on Solvency II: Standard formula SCR—Article 109 c life underwriting risk*. Committee of European Insurance and Occupational Pensions Supervisors. https://register.eiopa.europa.eu/CEIOPS-Archive/Documents/Advices/CEIOPS-L2-Final-Advice-on-Standard-Formula-Life-underwriting-risk.pdf

Cerchiara, R., Edwards, M., & Gambini, A. (2009). Generalized linear models in life insurance: Decrements and risk factor analysis under Solvency II. *Giornale dell'Istituto degli Attuari*, *72*, 100–122.

Chavez-Demoulin, V., & Davison, A. C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *54*(1), 207–222.

Chavez-Demoulin, V., Embrechts, P., & Hofert, M. (2016). An extreme value approach for modeling operational risk losses depending on covariates. *Journal of Risk and Insurance*, *83*(3), 735–776.

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer.

Consiglio, A., & De Giovanni, D. (2010). Pricing the option to surrender in incomplete markets. *Journal of Risk and Insurance*, *77*(4), 935–957.

Dar, A., & Dodds, C. (1989). Interest rates, the emergency fund hypothesis and saving through endowment policies: Some empirical evidence for the U.K. *Journal of Risk and Insurance*, *56*(3), 415–433.

DAV. (2018). *Herleitung der Sterbetafel DAV2008T für Lebensversicherungen mit Todesfallcharakter*. Deutsche Aktuarvereinigung. https://aktuar.de/unsere-themen/lebensversicherung/sterbetafeln/2018-10-05_DAV-Richtlinie_Herleitung_DAV2008T.pdf

Davison, A. C., & Huser, R. (2015). Statistics of extremes. *Annual Review of Statistics and its Application*, *2*(1), 203–235.

EIOPA. (2011). *EIOPA report on the fifth quantitative impact study (QIS5) for Solvency II*. European Insurance and Occupational Pensions Authority. https://register.eiopa.europa.eu/Publications/Reports/QIS5_Report_Final.pdf

EIOPA. (2019). *Risk-free interest rate term structures*. European Insurance and Occupational Pensions Authority. https://eiopa.europa.eu/tools-and-data/risk-free-interest-rate-term-structures-0_en

Eling, M., & Kiesenbauer, D. (2013). What policy features determine life insurance lapse? An analysis of the German market. *Journal of Risk and Insurance*, *81*(2), 241–269.

Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling extremal events*. Springer.

Embrechts, P., Mizgier, K., & Chen, X. (2018). Modeling operational risk depending on covariates: An empirical investigation. *Journal of Operational Risk*, *13*(3), 17–46.

Fang, H., & Kung, E. (2020). Why do life insurance policyholders lapse? The roles of income, health, and bequest motive shocks. *Journal of Risk and Insurance*, forthcoming.

Gottlieb, D., & Smetters, K. (2020). Lapse-based insurance. *American Economic Review*, forthcoming.

Hambuckers, J., Groll, A., & Kneib, T. (2018). Understanding the economic determinants of the severity of operational losses: A regularized generalized pareto regression approach. *Journal of Applied Econometrics*, *33*(6), 898–935.

Hambuckers, J., Kneib, T., Langrock, R., & Silbersdorff, A. (2018). A Markov-switching generalized additive model for compound Poisson processes, with applications to operational losses models. *Quantitative Finance*, *18*(10), 1679–1698.

Hitz, A., Davis, R., & Samorodnitsky, G. (2017). Discrete extremes. arXiv preprint 1707.05033.

Insurance Europe. (2019). *European insurance life industry database*. https://insuranceeurope.eu/sites/default/files/assets/DatabaseMarch2019_Life.xlsx

Kiesenbauer, D. (2012). Main determinants of lapse in the German life insurance industry. *North American Actuarial Journal*, *16*(1), 52–73.

Knoller, C., Kraut, G., & Schoenmaekers, P. (2016). On the propensity to surrender a variable annuity contract: An empirical analysis of dynamic policyholder behavior. *Journal of Risk and Insurance*, *83*(4), 979–1006.

Kuo, W., Tsai, C., & Chen, W.-K. (2003). An empirical study on the lapse rate: The cointegration approach. *Journal of Risk and Insurance*, *70*(3), 489–508.

Loisel, S., & Milhaud, X. (2011). From deterministic to stochastic surrender risk models: Impact of correlation crises on economic capital. *European Journal of Operational Research*, *214*(2), 348–357.

McNeil, A. J., Frey, R., & Embrechts, P. (2015). *Quantitative risk management: Concepts, techniques and tools*. Princeton University Press.

Milbrodt, H., & Helbig, M. (2008). *Mathematische Methoden der Personenversicherung*. Walter de Gruyter.

Milhaud, X., Loisel, S., & Maume-Deschamps, V. (2011). Surrender triggers in life insuance: What main features affect surrender behavior in a classical economic context? *Bulletin Français d'Actuariat*, *22*, 5–48.

Nadarajah, S., & Mitov, K. (2002). Asymptotics of maxima of discrete random variables. *Extremes*, *5*(3), 287–294.

Old Mutual. (2016). *Old mutual group's Solvency II and economic capital results*. http://otp.investis.com/clients/uk/old-mutual-plc/rns1/regulatory-story.aspx?cid=284&newsid=681804

Outreville, J. (1990). Whole life lapse rates and the emergency fund hypothesis. *Insurance: Mathematics and Economics*, *9*(4), 249–255.

Pesando, J. (1974). The interest sensibility of the flow of fund through life insurance companies: An economic analysis. *The Journal of Finance*, *29*(4), 1105–1121.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Probability*, *3*(1), 119–131.

Protektor. (2004). *Geschäftsbericht 2003*. https://protektor-ag.de/de/wp-content/uploads/sites/2/2015/07/protektor_geschaeftsbericht_2003.pdf

Renshaw, A., & Haberman, S. (1986). Statistical analysis of life assurance lapses. *Journal of the Institute of Actuaries*, *113*(3), 459–497.

Schott, F. (1971). Disintermediation through policy loans at life insurance companies. *The Journal of Finance*, *26*(3), 719–729.

SOA & LIMRA. (2012). *U.S. individual life insurance persistency*. Society of Actuaries and Life Insurance Marketing and Research Association. https://www.soa.org/globalassets/assets/files/research/exp-study/research-2007-2009-us-ind-life-pers-report.pdf

SOEP. (2019). German socio-economic panel, data for years 1984–2017, version 34. https://doi.org/10.5684/soep.v34

Standard & Poors. (2018). *European life insurers are playing the long game with product shifts*. https://allnews.ch/sites/default/files/files/European%20Life%20Insurers_22%20Feb%202018.pdf

UNIQA. (2017). *Group economic capital report*. http://uniqagroup.com/gruppe/versicherung/media/files/UNIQA_ECR_Report_2017_12042018.pdf

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Biagini F, Huber T, Jaspersen JG, Mazzon A. Estimating extreme cancellation rates in life insurance. *Journal Risk and Insurance*. 2021;1–30. https://doi.org/10.1111/jori.12336