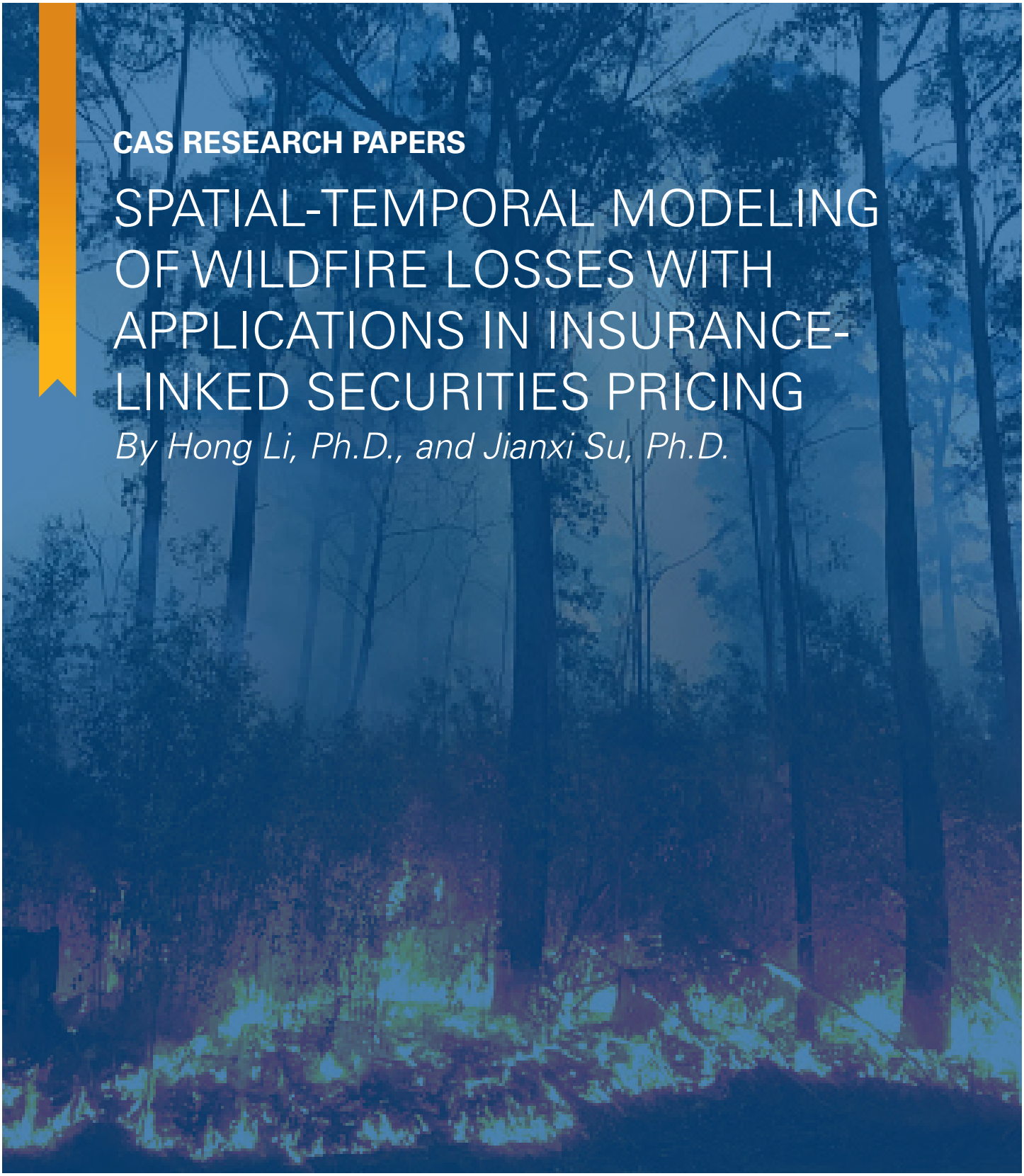


CAS RESEARCH PAPERS

SPATIAL-TEMPORAL MODELING OF WILDFIRE LOSSES WITH APPLICATIONS IN INSURANCE-LINKED SECURITIES PRICING

By Hong Li, Ph.D., and Jianxi Su, Ph.D.

CASUALTY ACTUARIAL SOCIETY

Spatial-Temporal Modeling of Wildfire Losses with Applications in Insurance-Linked Securities Pricing

Hong Li and Jianxi Su

Abstract: In this project, we model and predict state-specific wildfire losses in the United States using a combination of Bayesian dynamic models. In particular, the wildfire frequencies are modeled by a Bayesian multi-scale Dynamic Count Mixture Model (DCMM), which is capable of capturing a number of stylized features of wildfire data, including zero-inflation, over-dispersion compared to the Poisson distribution, and the time-varying patterns. Further, the DCMM is able to incorporate spatial dependence of different states, and thus improves the forecasting performance for individual states, especially those with low historical frequencies. Then we apply the predictive distribution of future wildfire loss to price wildfire catastrophe (CAT) bonds with different characteristics, and evaluate their hedging effectiveness for insurers in different states. We find that although using CAT bonds as a hedging tool may slightly increase the expected liability of an insurance portfolio due to bond premiums, the strategy can substantially reduce the variability risk and tail risk. Therefore, we conclude that CAT bond is a valuable tool of risk mitigation for insurers. Finally, for index-based CAT bonds whose payoffs are linked to wildfire losses in a larger area than that the insurer operates in, their hedging efficiencies are still acceptable. Therefore, it may be beneficial for insurers, especially those operates in areas with less frequent yet more volatile wildfire losses, to issue index-based CAT bonds, which are likely to be less expensive but much more liquid than indemnity bonds written directly on their liabilities.

Key words and phrases: Bayesian dynamic model; mixture model; counting process; wildfire loss; catastrophe bond

1. Introduction

In the wake of climate changes and human activities, wildfires have caused mounting damages over recent years, especially among the western states in the United States. Fifteen billion-dollar wildfire disasters have occurred in the U.S. since 2000. The financial losses are mainly caused by damage to homes and infrastructure. Beyond direct financial losses, wildfires further trigger indirect costs that can affect federal and state budget, public health, and natural environment. In 2018, wildfires in California alone had resulted in about $\$150$ billion (direct and indirect) losses, accounting for roughly 1.5% of California's annual gross domestic product (Wang, et al. 2021). Nationwide, the annualized economic burden due to wildfires is estimated to be as large as $\$350$ billion (Thomas, et al. 2017).

Insurance has naturally emerged as a financial tool for mitigating and managing wildfire risks. Insurance provides not only protection from economic hardships for property owners, but also incentives for *ex ante* risk reduction activities. Nevertheless, due to limited market capacity in the private insurance sector, the continually growing wildfire losses may significantly impact the availability of wildfire insurance. Another effective strategy for mitigating wildfire losses is through such innovative financial instruments as catastrophe (CAT) bonds. Multi-peril CAT bonds can be used to finance wildfire losses, but they may leave a significant degree of basis risk to insurers, that is, the payoffs received by the insurer may not be perfectly correlated with their actual losses due to wildfires. Wildfire-specific CAT bonds did not exist in the capital market until recently. Two pure California wildfire CAT bonds were issued in 2020: SD Re Ltd. (Series 2020-1) with a size of $90 million and Power Protective Re Ltd. (Series 2020-1) with a size of $50 million.[1] However, a comprehensive framework for modeling and pricing wildfire risk remains untouched in the actuarial literature, which may undermine the applications of such insurance-linked instruments for managing wildfire losses.

The purpose of this paper is two-fold. Firstly, we study the wildfire losses across different contiguous states of the U.S. over time by deploying an innovative Bayesian framework constituted by multi-variate count-valued time series processes and dynamic generalized linear models. The suggested Bayesian framework can incorporate spatial correlations of both the wildfire frequency and severity among different states, and accommodate the stylized facts that wildfire data are zero-inflated and heavy-tailed. Another advantage of using the proposed Bayesian framework is that, with a reasonable choice of prior distribution, the associated predictive distribution will be robust to outliers when data are scarce. It is also noteworthy that Bayesian framework has a dynamic nature, and the predictive distribution will be updated as new observations arrive, suitable for the purposes of continuous risk monitoring and dynamic risk pricing.

[1] Source: https://www.artemis.bm/deal-directory/sd-re-ltd-series-2020-1/ and https://www.artemis.bm/deal-directory/power-protective-re-ltd-series-2020-1/

Secondly, capitalized on the aforementioned wildfire loss models, we will propose a sophisticated yet transparent framework for pricing wildfire insurance-linked securities. In particular, we have special interests in studying the following fundamental questions in wildfire risk management:

1. How would the prices of wildfire-linked CAT bonds vary with the bond characteristics, such as the loss trigger and the exhaustion point?
2. What is the basis risk inherent in index-based CAT bonds?
3. What is the impact of spatial dependence on pricing wildfire CAT bonds?

Bearing the two aforementioned practical goals in mind, the rest of this paper will proceed as follows. Section 2 contains a description of the wildfire data we are going to use in this current paper, which are retrieved from the Spatial Hazard Events and Losses Database (SHELDUS).[2] An innovative and sophisticated modeling framework will be proposed in Section 3 to capture the unique features exhibited in the SHELDUS wildfire data. Section 4 summarizes the implementation of the proposed framework and outlines the results. Applications of the model's results in studying CAT bonds will be examined in Section 5. Then in Section 6, we extend the applications of our models to a reinsurance analysis. Section 7 focuses on modeling and predicting the county-level wildfire losses in California where the occurrences of wildfires are very frequent. Section 8 contains our conclusions.

2. A description of the wildfire data in SHELDUS

SHELDUS is a comprehensive hazard database maintained by the Center for Emergency Management and Homeland Security at Arizona State University. The latest version of SHELDUS data tracks the occurrence and loss records of numerous perils—including wildfires in particular—occurred across the U.S. from 1960 to 2019. The wildfire data in SHELDUS are compiled from the event records of the National Centers for Environmental Information (NOAA).[3] Compared to other wildfire data sets commonly used for environmental modeling purposes, the SHELDUS database is more suitable for insurance studies because it contains the direct financial loss associated with the wildfire events.

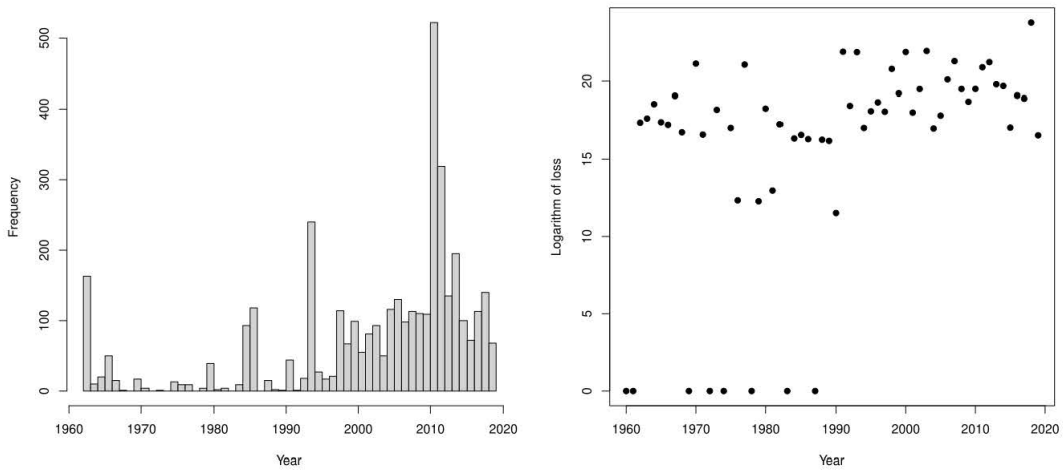
In the SHELDUS data, each hazard record consists of environmental information such as disaster occurrence month, year, state, and duration, as well as damage information such as financial losses on crops and properties, human injuries, and fatalities. Figure 1 displays the nation-wise aggregate wildfire frequency and the crop and property losses per year recorded in the SHELDUS data. For a fair comparison, financial losses are adjusted according to the Consumer Price Index with the year 2019 being the base period. We observe that the annual number of wildfire occurrences across the country started to grow dramatically since the mid-1990s and reached a peak around 2010. Although the number of wildfires seemed to decrease over the past few years, it is still very high compared to

[2] Source: https://cemhs.asu.edu/sheldus

[3] Source https://www.ncdc.noaa.gov/IPS/sd/sd.html.

the period before 2000. Consequently, there is also an apparent increase in annual aggregate wildfire losses over time, and an extremal wildfire loss record is observed in 2018. The wildfire loss pattern observed generally coincides with the one reported from the insurance industry (Webb and Xu 2018).

Figure 1. The nation-wise total number of wildfire occurrences per year (left) and the natural logarithm of annual aggregate loss (right) from 1960 to 2019.



Another important feature of wildfires in the U.S. is that their occurrence frequencies vary significantly across different territories. Table 1 summarizes the decennial wildfire frequencies among the 45 unique states reported in the SHELDUS data. The top six states having the greatest number of wildfires occurring from 1960 to 2019 are California, Colorado, Montana, New Mexico, Texas, and Washington. It is noted that California consistently had the greatest number of wildfires among different year intervals, yet wildfires became noticeably more frequent in Colorado during the past two decades. In Montana, there was a peak number of 120 wildfires in 1990s, but the number seemed to decline over recent years. In Texas and Washington, a significant growth in wildfire occurrence was observed in the 2010s, with the numbers approximately tripled the ones in 2000s.

We note that a few outliers appear in Table 1. For instance, there were 100 wildfires in North Carolina in the 1980s. This is due to a massive wildfire that started off on May 5, 1986, in the Pender County. The fire lasted for 11 days and grew to 73,000 acres, spreading the whole state, before it was controlled.[4] 100 counties in the state were affected and reported the fires separately. SHELDUS recorded the occurrence of wildfires at the county level, so there were 100 wildfire records for North Carolina in 1986. In SHELDUS data, we also observed that South Carolina recorded the most wildfires in a single month in March 1985.[5] The number of wildfires was also quite large in the following month. Moreover, a series of

[4] Source: https://www.wect.com/story/31902417/massive-pender-county-wildfire-started-30-years-ago/.

[5] Source: https://www.state.sc.us/forest/firesign.htm.

destructive wildfires occurred in Texas in 2011. In particular, 47.3% of all acreage burned in the U.S. in 2011 was burned in Texas. The fires had been particularly severe due to the 2011 Southern U.S. drought that covered the state, and was exacerbated by the unusual convergence of strong winds, unseasonably warm temperatures, and low humidity.[6]

A summary of the decennial wildfire frequencies among the states recorded in the SHELDUS data, and the shaded entries represent the top six states having most wildfire occurrences.

Figure 2 displays the annual wildfire frequencies as well as the corresponding crop and property losses for the top six most wildfire states outlined above. In general, the occurrences of wildfires are highly volatile over time with rather complex dependencies existing among different states. Therefore, flexible modeling framework must be adopted in order to obtain satisfactory predictions of future wildfire losses.

3. Spatial-Temporal modeling of aggregate wildfire losses

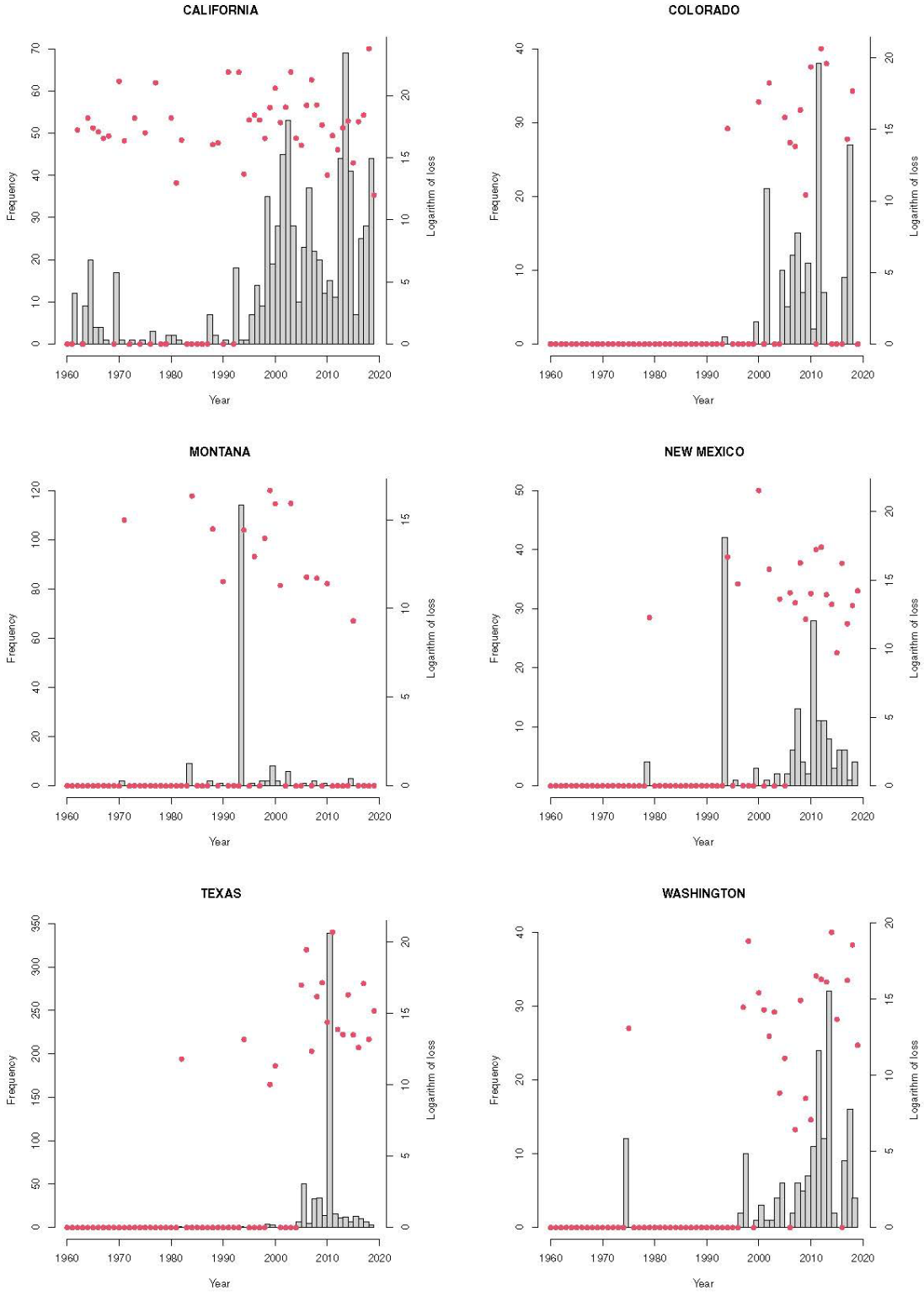
In this section, we aim to put forth a modeling framework based on Bayesian dynamic state-space models (West and Harrison 2006) for predicting state-specific wildfire losses. It is noted that separate frequency and severity modeling is commonly viewed as the best practice in the context of property and casualty insurance (Klugman, Panjer and Willmot 2012). Thus we generalize the classical Bayesian state-space models with the well-known "two-part" structure (Frees 2009). Specifically, the wildfire frequency will be studied using a novel Bayesian Dynamic Count Mixture Model (DCMM) proposed in (Berry and West 2020), and the loss severity will be fitted by a Bayesian Dynamic Linear Model (DLM). The class of predictive Bayesian state-space models have been shown to be useful in a wide range of time-series contexts. In the context of wildfire modeling particularly, the merits of adopting the Bayesian state-space models include the following. First, Bayesian state-space models can cope with many nonstandard data features such as zero-inflation and over-dispersion in wildfire frequency and heavy-tailed pattern in wildfire severity. Second, the embedded Bayesian approach implicitly captures the uncertainties surrounding the point estimates. As such, interval prediction of future wildfire losses can be constructed. Third, Bayesian methods are more suitable for estimating models with complex structures, such as multi-level regression models (McElreath 2018) and the state-space models presented in this paper. The multi-step predictive distribution of wildfire losses can be simulated efficiently via the state-space updating algorithm, without resorting to MCMC or other computationally intensive algorithms.

[6] Source: https://en.wikipedia.org/wiki/2011_Texas_wildfires.

Table 1. A summary of the decennial wildfire frequencies among the states recorded in the SHELDUS data; shaded entries represent the top six states having most wildfire occurrences.

State	1960–1969	1970–1979	1980–1989	1990–1999	2000–2009	2010–2019	Total	Ranking
Alabama	0	0	0	0	0	2	2	42
Alaska	0	6	0	4	34	9	53	19
Arizona	0	0	0	0	11	34	45	23
Arkansas	0	0	0	0	5	25	30	28
California	50	23	14	86	285	296	754	1
Colorado	0	0	0	1	73	94	168	4
Connecticut	1	0	0	0	0	0	1	44
Delaware	8	0	0	0	2	0	10	36
Florida	0	0	0	95	33	20	148	8
Georgia	0	0	0	0	31	108	139	9
Guam	0	0	0	0	1	0	1	44
Hawaii	0	0	0	0	12	11	23	30
Idaho	10	0	0	15	17	84	126	11
Illinois	0	0	0	0	2	4	6	40
Iowa	0	0	0	0	1	5	6	40
Kansas	0	0	0	2	3	16	21	31
Kentucky	0	0	0	13	0	2	15	34
Louisiana	0	0	0	0	0	21	21	31
Maryland	48	0	0	0	1	0	49	21
Michigan	0	0	0	3	9	5	17	33
Minnesota	0	1	36	0	4	6	47	22
Missouri	0	0	0	0	18	19	37	24
Montana	0	2	11	120	19	4	156	6
Nebraska	0	0	0	1	1	26	28	29
Nevada	0	0	0	0	0	33	33	25
New Jersey	33	0	0	4	4	10	51	20
New Mexico	0	4	0	43	31	80	158	5
New York	63	0	1	1	4	0	69	16
North Carolina	0	0	100	0	1	9	110	13
North Dakota	0	0	0	0	4	7	11	35
Oklahoma	0	0	0	0	38	39	77	15
Oregon	0	0	0	3	35	22	60	17
Pennsylvania	0	0	0	0	4	5	9	37
Puerto Rico	0	0	0	0	54	64	118	12
South Carolina	46	0	99	0	1	6	152	7
South Dakota	0	0	0	7	2	0	9	37
Tennessee	0	0	0	0	0	2	2	42
Texas	0	0	1	5	132	432	570	2
Utah	0	0	0	1	6	123	130	10
Virgin Islands	0	0	0	0	7	0	7	39
Virginia	0	0	2	16	9	5	32	27
Washington	0	12	0	12	29	117	170	3
West Virginia	0	9	0	45	1	2	57	18
Wisconsin	0	0	18	73	7	7	105	14
Wyoming	0	0	0	0	14	19	33	25

Figure 2. The annual wildfire frequencies (in bar plots) as well as the corresponding crop and property losses (in dot plots) for the top six most wildfire states: California, Colorado, Georgia, New Mexico, Texas, and Washington.



3.1 The single-scale DCMM frequency model

To facilitate the introduction of the proposed modeling framework, in this part of the paper we should focus on state-level, monthly wildfire data. However, the proposed framework can be applied to study any data with different levels of granularity in space and time. The Bayesian DCMM for frequency modeling is first introduced. Let us begin with the so-called single-scale models in which the wildfire frequencies of individual states are model independently, without utilizing information from the other states. In the actuarial literature, Poisson processes have been a prevalent model choice for studying the occurrence of insurance events. When it comes to modeling wildfires, some territories may have high probabilities of zero wildfire event over time. Classical Poisson models must be modified in order to account for the highly zero-inflated structure presented in wildfire data. Suppose that there are in total N states and T months in the sample, and let $n_{i,t}$ be the number of wildfires occurred in state i at time t. The wildfire frequency is assumed to follow a mixture of Bernoulli and (over-dispersed) Poisson distributions. Formally, define the binary series $z_{i,t} = \mathbf{1}(n_{i,t} > 0)$, where $\mathbf{1}(\cdot)$ is the indicator function. The distributions of $z_{i,t}$ is given by:

$$z_{i,t} \sim \text{Ber}(\pi_{i,t}),$$

where $\text{Ber}(\pi_{i,t})$ denotes the Bernoulli distribution with $\pi_{i,t}$ being the probability that $z_{i,t} = 1$, i.e., $\mathbb{P}(z_{i,t} = 1) = \pi_{i,t}$. Conditional on the Bernoulli variable $z_{i,t}$, the distribution of $n_{i,t}$ is then specified as

$$n_{i,t}|z_{i,t} = \begin{cases} 0, & \text{if } z_{i,t} = 0, \\ 1 + x_{i,t}, \quad x_{i,t} \sim \text{Poi}(\mu_{i,t}), & \text{if } z_{i,t} > 0, \end{cases}$$

where $\text{Poi}(\mu_{i,t})$ denotes the Poisson distribution with mean $\mu_{i,t}$. In the model above, the Bernoulli variable $z_{i,t}$ characterizes the probability of not having any wildfire in state i at time t, and given the occurrence of wildfire, the shifted Poisson variable $n_{i,t}$ captures the distribution of wildfire frequency.

The parameters in the two distributions above, $\pi_{i,t}$ and $\mu_{i,t}$, are fitted separately by a transformed linear state-space model:

$$\begin{cases} logit(\pi_{i,t}) = \boldsymbol{F}_{i,t,\pi}^{\top} \boldsymbol{\xi}_{i,t}, & with\ \boldsymbol{\xi}_{i,t} = \Gamma_{i,t,\pi}\ \boldsymbol{\xi}_{i,t-1} + \boldsymbol{\omega}_{i,t}, \\ log(\mu_{i,t}) = \boldsymbol{F}_{i,t,\mu}^{\top}\ \boldsymbol{\theta}_{i,t}, & with\ \boldsymbol{\theta}_{i,t} = \Gamma_{i,t,\mu}\ \boldsymbol{\theta}_{i,t-1} + \boldsymbol{\eta}_{i,t}. \end{cases} \tag{1}$$

Equation (1) includes a pair of generalized linear models (GLM). First, the logit function $\text{logit}(\cdot)$ maps the probability $\pi_{i,t} \in (0,1)$ to the real line $(-\infty, \infty)$, and the log function $\log(\cdot)$ maps the (positive) Poisson mean $\mu_{i,t} \in (0, \infty)$ to the real line $(-\infty, \infty)$. These two functions are commonly used *link functions* in the context of GLM, which connect bounded variables

of interest, $\pi_{i,t}$ and $\mu_{i,t}$, to the linear regression components, $\mathbf{F}_{i,t,\pi}^{\top}\ \boldsymbol{\xi}_{i,t}$ and $\mathbf{F}_{i,t,\mu}^{\top}\ \boldsymbol{\theta}_{i,t}$, respectively. In the above formulas, $\mathbf{F}_{i,t,\pi}$ and $\mathbf{F}_{i,t,\mu}$ denote vectors of known values of predictor variables, $\boldsymbol{\xi}_{i,t}$ and $\boldsymbol{\theta}_{i,t}$ are time-varying latent state vectors for capturing the dynamics of $z_{i,t}$ and $n_{i,t}$, respectively, and $\Gamma_{i,t,\pi}$ and $\Gamma_{i,t,\mu}$ are evolution matrices for specifying structural changes of the state vectors over time, to be specified by the model user depending on the modeling task at hand. When the data have no significant trend, seasonality or cyclic behavior, then evolution matrices $\Gamma_{i,t,\pi}$ and $\Gamma_{i,t,\mu}$ can be as simple as identity matrices, and the corresponding state vectors are simply random walk processes. Otherwise, appropriate functional forms for the evolution matrices will be chosen to capture the time series pattern presented in data. Finally, $\boldsymbol{\omega}_{i,t}$ and $\boldsymbol{\eta}_{i,t}$ are independent stochastic vectors of error terms with (conditional) zero means and covariance matrices $\mathbf{K}_{i,\omega,t}$, $\mathbf{K}_{i,\eta,t}$, respectively.

The proposed DCMM is capable of handling rather intricate wildfire dynamics. For instance, the embedded Bernoulli component accommodates the zero-inflation issue encountered in wildfire data, and the dynamic nature of $\pi_{i,t}$ and $\mu_{i,t}$ can address the time-varying variability in wildfire frequency. If the observed wildfire data are over-dispersed relative to the conditional Poisson model, a random effect can be further added to the Poisson state-space model to make the forecast distributions more accurate in terms of capturing extremal observations. Specifically, the Poisson state-space model is extended to

$$\log\left(\mu_{i,t}\right) = \mathbf{F}_{i,t,\mu}^{\top}\ \boldsymbol{\theta}_{i,t} + r_{i,t}, \quad \text{with } \boldsymbol{\theta}_{i,t} = \Gamma_{i,t,\mu}\ \boldsymbol{\theta}_{i,t-1} + \boldsymbol{\eta}_{i,t}, \tag{2}$$

where $r_{i,t}$ is a time-t specific, independent, zero-mean random effect for state i. The extended Poisson model (2) recognizes the time-t individual and unpredictable variation above the time-dependent component $\mathbf{F}_{i,t,\mu}^{\top}\ \boldsymbol{\theta}_{i,t}$.

3.2 The multi-scale DCMM frequency model

Several empirical studies have suggested that the occurrences of wildfires are closely related to climate change and human activities, thus the wildfire frequencies across different states may share some common patterns over time. Taking into account the dependencies of wildfire occurrences among different states might help improve the modeling and predictive accuracy of wildfire frequencies for individual states. To this end, a set of common factors $\boldsymbol{\phi}_t$ are introduced into the feature vectors $\mathbf{F}_{i,t,\pi}$ in Equation (1) so as to account for the common drivers of wildfires such as seasonal effects and other stochastic effects not explained by the state-specific covariates. Now the feature vectors in (1) become

$$\mathbf{F}_{i,t,\pi} = \begin{pmatrix}\mathbf{f}_{i,t,\pi} & \boldsymbol{\phi}_t\end{pmatrix}^{\top} \quad \text{and} \quad \mathbf{F}_{i,t,\mu} = \begin{pmatrix}\mathbf{f}_{i,t,\mu} & \boldsymbol{\phi}_t\end{pmatrix}^{\top}, \tag{3}$$

where $\mathbf{f}_{i,t,\pi}$ and $\mathbf{f}_{i,t,\mu}$ contain respectively the state-specific covariates in the Bernoulli and Poisson regressions, and $\boldsymbol{\phi}_t$ contains the common factors. The impact of these common factors are specific to each state, captured by the state vectors $\boldsymbol{\xi}_{i,t}$ and $\boldsymbol{\theta}_{i,t}$. To summarize, the multi-scale DCMM frequency model considered in this project is formulated as

$$\begin{cases} \text{logit}(\pi_{i,t}) = \mathbf{F}_{i,t,\pi}^{\top}\ \boldsymbol{\xi}_{i,t}, & \text{with } \boldsymbol{\xi}_{i,t} = \Gamma_{i,t,\pi}\ \boldsymbol{\xi}_{i,t-1} + \boldsymbol{\omega}_{i,t}, \\ \log(\mu_{i,t}) = \mathbf{F}_{i,t,\mu}^{\top}\ \boldsymbol{\theta}_{i,t} + r_{i,t}, & \text{with } \boldsymbol{\theta}_{i,t} = \Gamma_{i,t,\mu}\ \boldsymbol{\theta}_{i,t-1} + \boldsymbol{\eta}_{i,t}, \end{cases} \tag{4}$$

with the regressors $\mathbf{F}_{i,t,\pi}$ and $\mathbf{F}_{i,t,\mu}$ given in Equation (3).

In principle, the common factors $\boldsymbol{\phi}_t$ could be latent and are implied jointly by the models (4) of all states. Nevertheless, simultaneously estimating all state-specific models is computationally onerous, or even prohibitive if the dimension of $\boldsymbol{\phi}_t$ is large. Further, estimating the common latent factors itself is a very challenging problem. As a remedy of this computation issue, (Berry and West 2020) proposed to approximate the latent factors by a set of *observable* factors. For example, $\boldsymbol{\phi}_t$ could be variables representing some common patterns of the whole system. In our case, $\boldsymbol{\phi}_t$ could be the observed meteorological variables of the whole nation or a geographical region consisting of multiple states. Alternatively, $\boldsymbol{\phi}_t$ could be some systematic factors extracted from the national or regional wildfire frequency data. As will be detailed in Section 4, in this research we will let $\boldsymbol{\phi}_t$ be the seasonal effects extracted from the regional wildfire frequencies.

After specifying the common factors $\boldsymbol{\phi}_t$, they are simply treated as observable variables in the state-specific models. As a result, one can estimate Model (4) for each state separately. This greatly reduces the computational burden of the analysis, and allows us to increase the number of local models conveniently. For instance, one can jointly model the wildfire loss of counties instead of states, when related data are available. Finally, the Bayesian state space algorithm is used to estimate the proposed model. In a nutshell, with the (user-specified) initial prior distribution of $\boldsymbol{\xi}_{i,0}$, $\boldsymbol{\theta}_{i,0}$, and $\boldsymbol{\phi}_0$, the posterior distributions of these parameters are updated recursively when $z_{i,t}$ and $n_{i,t}$ are observed for each i and t. The predictive distributions of $z_{i,T+k}$ and $n_{i,T+k}$ can then be calculated (via simulation) for each state i and forecast horizon k. For details about the estimation algorithm, we refer to (Berry and West 2020) and Chapter 4 of (Prado and West 2010).

3.3 The DLM loss severity model

We propose to use a Bayesian DLM to model wildfire losses. Specifically, let $y_{i,t}$ be the *average* wildfire claim amount for state i at time t and $\mathbf{F}_{i,t,y}$ be the corresponding covariate vector, the severity processes are given by:

$$g(y_{i,t}) = \mathbf{F}_{i,t,y}^{\top}\ \boldsymbol{\beta}_{i,t}, \quad \text{with} \quad \boldsymbol{\beta}_{i,t} = \boldsymbol{\Gamma}_{i,t,\beta}\ \boldsymbol{\beta}_{i,t-1} + \boldsymbol{\varepsilon}_{i,t}, \quad i = 1, \ldots, N,$$

where $g(\cdot)$ is a link function (such as the log-link function). The DLM loss severity model will be estimated using a Bayesian state space algorithm similar to the DCMM frequency model.

4. Implementation and results

In this section we apply the models introduced in Section 3 to study the SHELDUS wildfire data. As we have seen in Table 1, some states have only few wildfires over the whole sample. Therefore, we leave out the states with 15 or fewer wildfires shown in Table 1, as these states have small risk exposure to wildfire, and might cause numerical problems in the estimation procedure due to excessive zeros in their samples. We also leave out Puerto Rico as it is a Caribbean island and unincorporated U.S. territory so both its geographical and socioeconomic conditions may be rather different from the other states. Further, as can be seen in Figure 2, for many states there have been only few records (or even no record) of wildfire over the first few decades. Therefore, the first few decades of the sample does not seem to be relevant in predicting future wildfire occurrence, and to make things worse, numerical issues may arise when (almost) all data points are zero. Therefore, in this research we consider the data from 1989 onward, and so each state is left with 372 observations (January 1989 to December 2019). The DCMM frequency model is implemented via the Python package *PyBATS,*[7] and the DLM loss severity model is implemented via the Python package *PyDLM.*[8]

4.1 Data pre-processing and set-up

After cleaning the data, we are left with 32 states. Before the analysis, we further divide the remaining states into two groups based on their geographic locations: the Western region (including the West and Southwest regions)[9] and the Eastern region (including the Northeast, Southeast, and Midwest regions). The Western (resp. Eastern) Region consists of 15 (resp. 17) states. The compositions of the two groups are summarized in Table 2.

Table 2. States in the Western and Eastern regions.

Western Region				Eastern Region			
Alaska	Arizona	California	Colorado	Arkansas	Florida	Georgia	Kansas
Hawaii	Idaho	Montana	Nevada	Louisiana	Maryland	Michigan	Minnesota
New Mexico	Oklahoma	Oregon	Texas	Missouri	Nebraska	New Jersey	New York
Utah	Washington	Wyoming		North Carolina	South Carolina	Virginia	West Virginia
				Wisconsin			

[7] Source: https://lavinei.github.io/pybats.

[8] Source: https://pydlm.github.io/pydlm_user_guide.html.

[9] The definition of major regions in the United States can be found on https://www.ducksters.com/geography/us_states/us_geographical_regions.php.

The empirical application proceeds as follows:

1. We construct the two region level time series data by aggregating the corresponding state level data.

2. For each region, we extract the common factors from the regional data. At this step, we apply a separate Bayesian Dynamic Generalized Linear model (DGLM) to the regional data, and extract the first 6 harmonic seasonal factors from the DGLM model. Intuitively speaking, harmonic functions consist of Fourier functions which represent the periodic fluctuations. The linear combinations of the harmonic functions are then used to represent the seasonal effects embedded in the data. For detailed discussions of harmonic factors and the Fourier representation of seasonality, we refer to (West and Harrison 2006).

3. After obtaining the harmonic seasonal factors, we treat them as *observed* and use them as known regressors in the corresponding state-specific models (4) for predicting state level wildfire frequencies.

For each state, we use the first 15 years of data to estimate the model. Then, we perform out-of-sample forecasts from 1 month to 12 months on a rolling window basis over the remaining 16 years (2004 to 2019) of data. Specifically, at the first step, we estimate the model using data from January 1989 to December 2003, and perform monthly out-of-sample forecasts of wildfire frequencies from January 2004 to December 2009. As the second step, we update the data by one month, estimate the model using data from January 1989 to January 2005, and perform monthly out-of-sample forecasts from February 2005 to January 2010. We continue this process until the final step, where we fit the model using data from January 1989 to December 2018, and perform monthly out-of-sample forecasts from January 2019 to December 2019.

Finally, in this project, we rely solely on the historical wildfire data to predict their future patterns. The reason we exclude exogenous covariates, such as climate data, is threefold. First, there are plenty of climate variables that could potentially affect the occurrence of wildfire, and thus should be considered as predictors. However, due to the sample size restriction, it is not possible to simultaneously include all these variables in the model. Therefore, appropriate variable selection and/or dimension reduction techniques should be applied to the (large) set of covariates, before the model can be estimated. Second, we found that the harmonic seasonal factors can already capture the wildfire patterns quite well, and the inclusion of climate variables (such as average or maximum/minimum of monthly temperature and precipitation) can only marginally improve, or sometimes even deteriorate, the forecasting performance of the model. In other words, the inclusion of exogenous climate variables could lead to over-fitting and introduce extra noise into the model. Third, in the presence of exogenous covariates, it is more difficulty to conduct forecasting analysis, because we have to "predict the predictors" before predicting the wildfire frequencies. This is a difficult task because climate variables themselves are hard to project. Nevertheless, we note that there are also advantages of including exogenous

variables. The most important advantage is probably that including these variables enables one to analyze their cross-sectional correlation with wildfire occurrence, and hence allows one to evaluate the impact of climate or socioeconomic activities on wildfire. Moreover, the inclusion of exogenous variables may be useful for insurance practices. For instance, various actuarial societies in North America have jointly proposed the Actuarial Climate Index (ACI)[10] recently. The ACIs are aggregate climate indices measured at 12 regions in North America from 1961 to 2020 at the monthly frequency. These indices are related to various insurance-related events, such as crop yields and natural perils and hazards, and could potentially be used to design and price weather-linked derivatives. Therefore, the inclusion of the ACI may shed light on the pricing of wildfire-linked CAT bond or (re)insurance practices on wildfire. We leave the inclusion of exogenous variables to future research.

4.2 Wildfire frequency forecasting

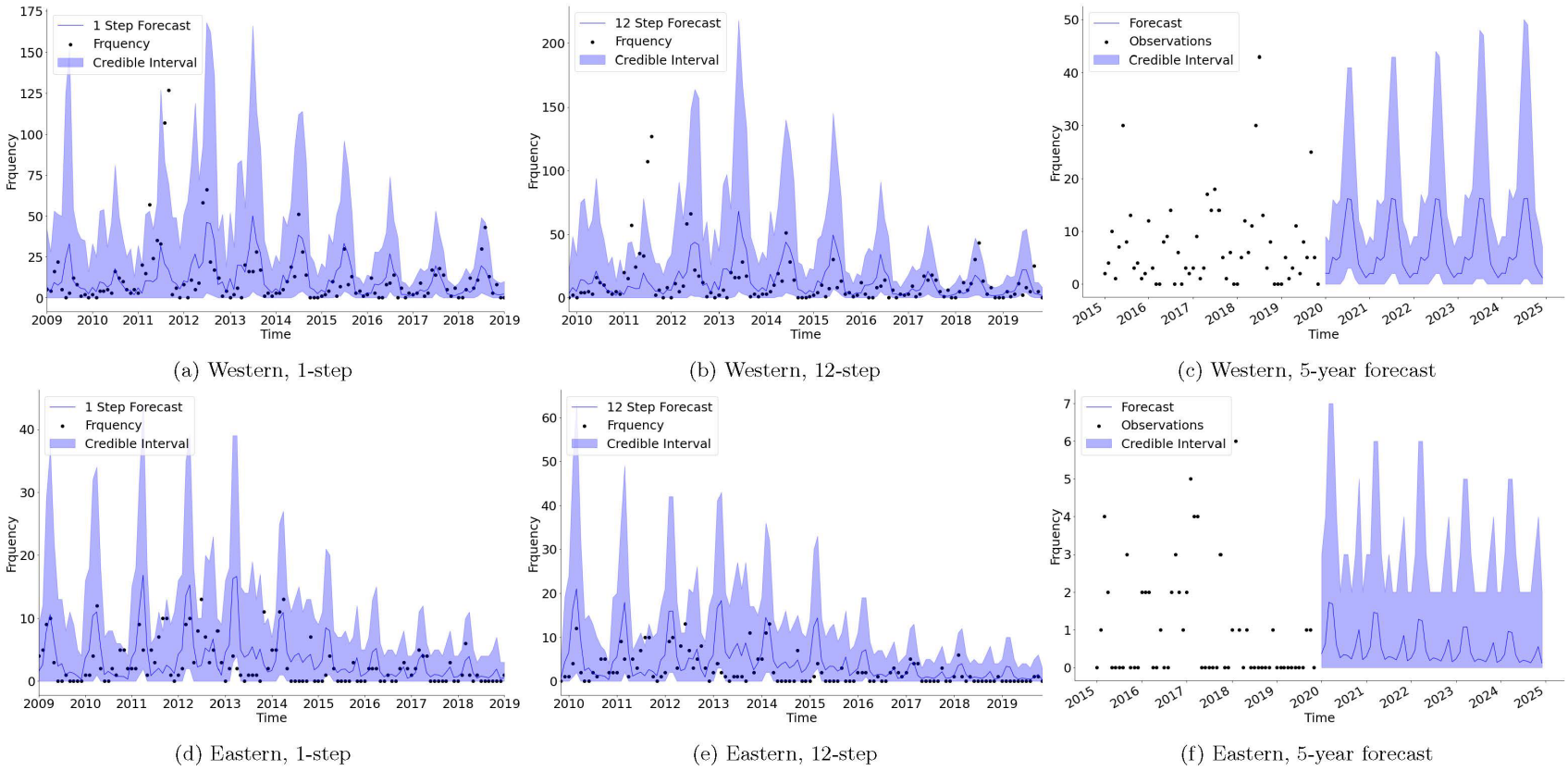
First, we display the 1-month and the 12-month out-of-sample forecast for the Western and the Eastern regions in Figure 3. Further, in order to examine the long-term forecast performance of the DCMM model, we show the 5-year forecast from January 2020 to December 2024 for both regions in the leftmost column of Figure 3. From the figure, we can see that much more wildfires (6.52 wildfires per month on average) have occurred in the Western region than the Eastern region (1.77 wildfire per month on average) over the sample. Hence, the Western region is more heavily exposed to wildfire risk. The high number of wildfire frequency also leads to smoother estimation of both the mean forecast and the 99% credible intervals for the Western region. Further, the occurrence of wildfire exhibits clear seasonal patterns, with most wildfires occurring in the summers and falls. More importantly, solely based on historical frequencies, the DCMM is able to capture the time-varying patterns of wildfire frequencies in both regions and the 99% credible intervals cover most of the observations for both the 1-month and 12-month forecasts. Furthermore, we can see that the model is learning from past data in generating future forecasts, thanks to its dynamic feature. For example, a larger observation will lead to both larger mean forecasts and wider credible intervals in the next period. The forecasting results are rather similar with other forecasting horizons. Finally, the 5-year forecasts of the two regions are rather different: The projected frequency of the Western region is increasing over time, while the one of the Eastern region is decreasing. The opposite forecasts reflect the diverging historical pattern embedded in the regional data.

Next, we look at the state-specific models. As illustrations, Figure 4 and Figure 6 display the 1-month and 12-month out-of-sample forecasts, as well as the long-term forecasts for California, Colorado, Oregon, and Texas from the Western region, and Florida and Georgia from the Eastern region. Among the states we select, California and Texas (resp. Florida and Georgia) are the top 2 states with the most wildfires over the sample of January 1989 to December 2019 in the Western (resp. Eastern) region. Colorado and Oregon are the

[10] Source: https://actuariesclimateindex.org/home/.

states of interest for existing wildfire CAT bonds. Again, we see that the model produces smoother forecasts and clearer seasonal patterns for the states with more historical wildfires, such as California and Texas. Moreover, the 99% credible intervals are able to cover most of the observations, especially for the 1-month forecast. One exception is Texas over 2011, where the wildfire frequency was extraordinarily high compared to the former and the later years. The DCMM model failed to capture this abnormality, as there were no similar historical observations to learn from. However, we see that the model responded by producing wider credible intervals in the following year, with the magnitude comparable to the outliers. In later years, the realized wildfire frequencies reduced to the average level, and thus the model predictions.

Figure 3. The 1-step (left) and 12-step (mid) mean out-of-sample forecast and the 99% credible intervals, and the 5-year forecasts into the future (right) for the Western region (upper panel) and the Eastern region (lower panel).



(a) Western, 1-step

(b) Western, 12-step

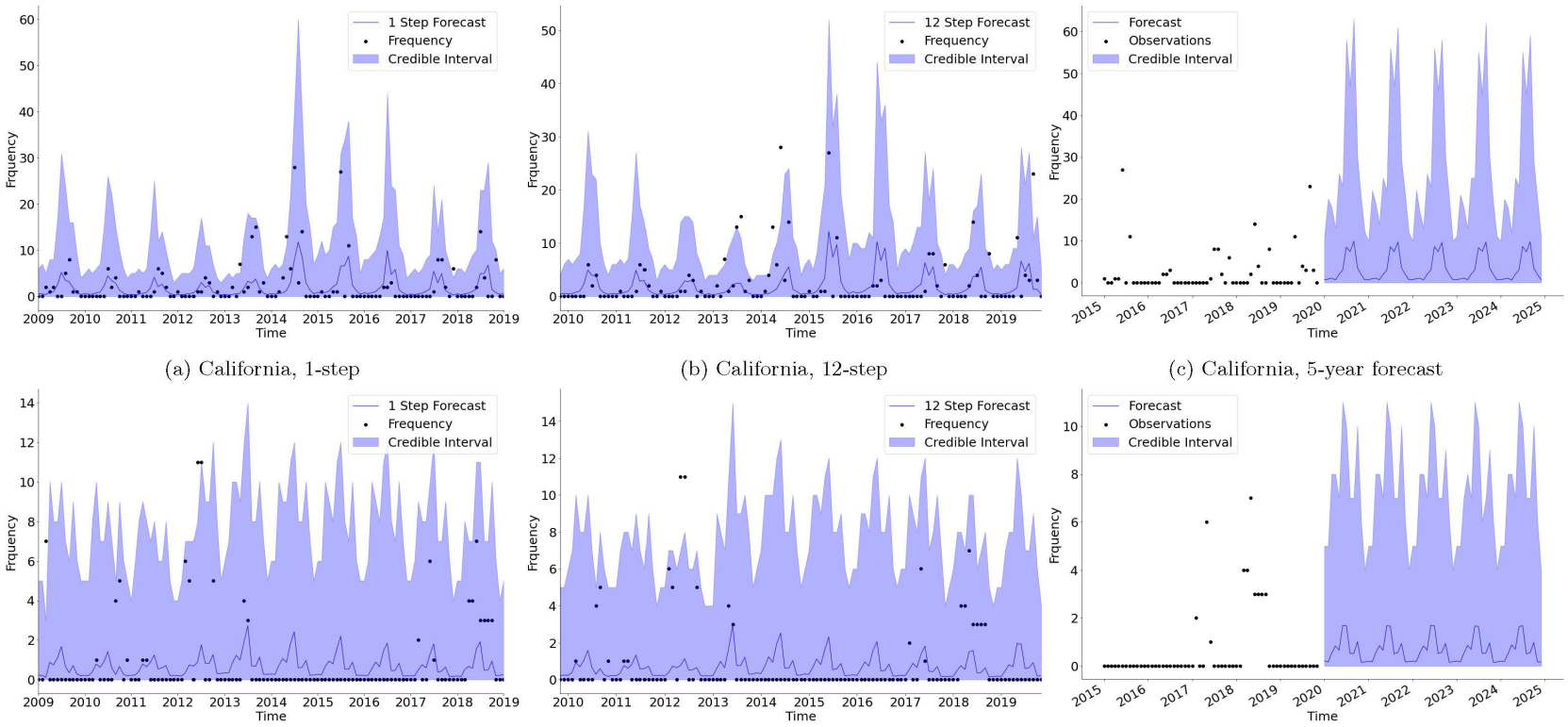
(c) Western, 5-year forecast

(d) Eastern, 1-step

(e) Eastern, 12-step

(f) Eastern, 5-year forecast

Figure 4. The 1-step (left) and 12-step (mid) mean out-of-sample forecast and the 99% credible intervals, and the 5-year forecasts into the future (right) for California (upper panel) and Colorado (lower panel).



(a) California, 1-step (b) California, 12-step (c) California, 5-year forecast

(d) Colorado, 1-step (e) Colorado, 12-step (f) Colorado, 5-year forecast

Figure 5. The 1-step (left) and 12-step (mid) mean out-of-sample forecast and the 99% credible intervals, and the 5-year forecasts into the future (right) for Oregon (upper panel) and Texas (lower panel).

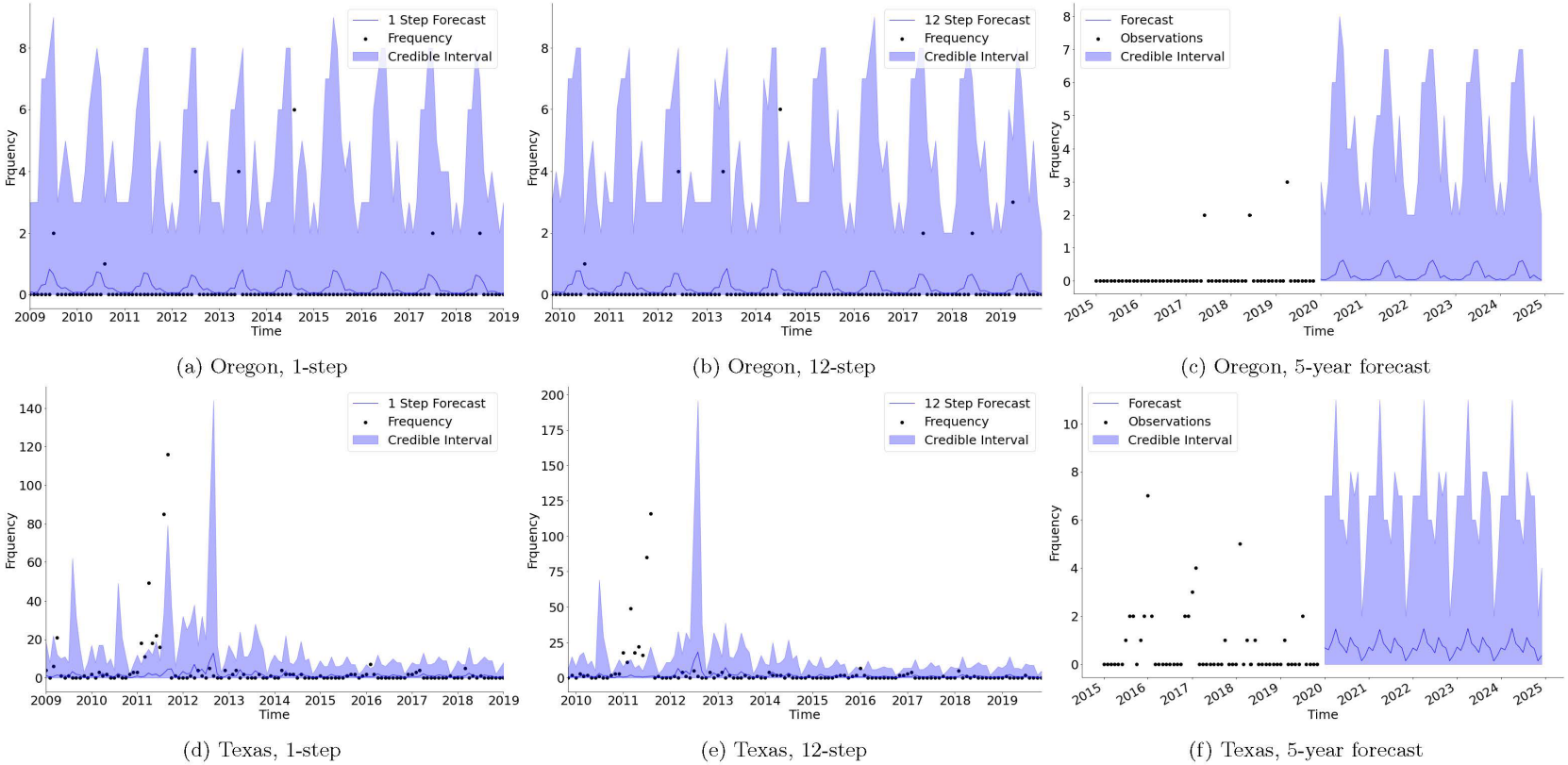
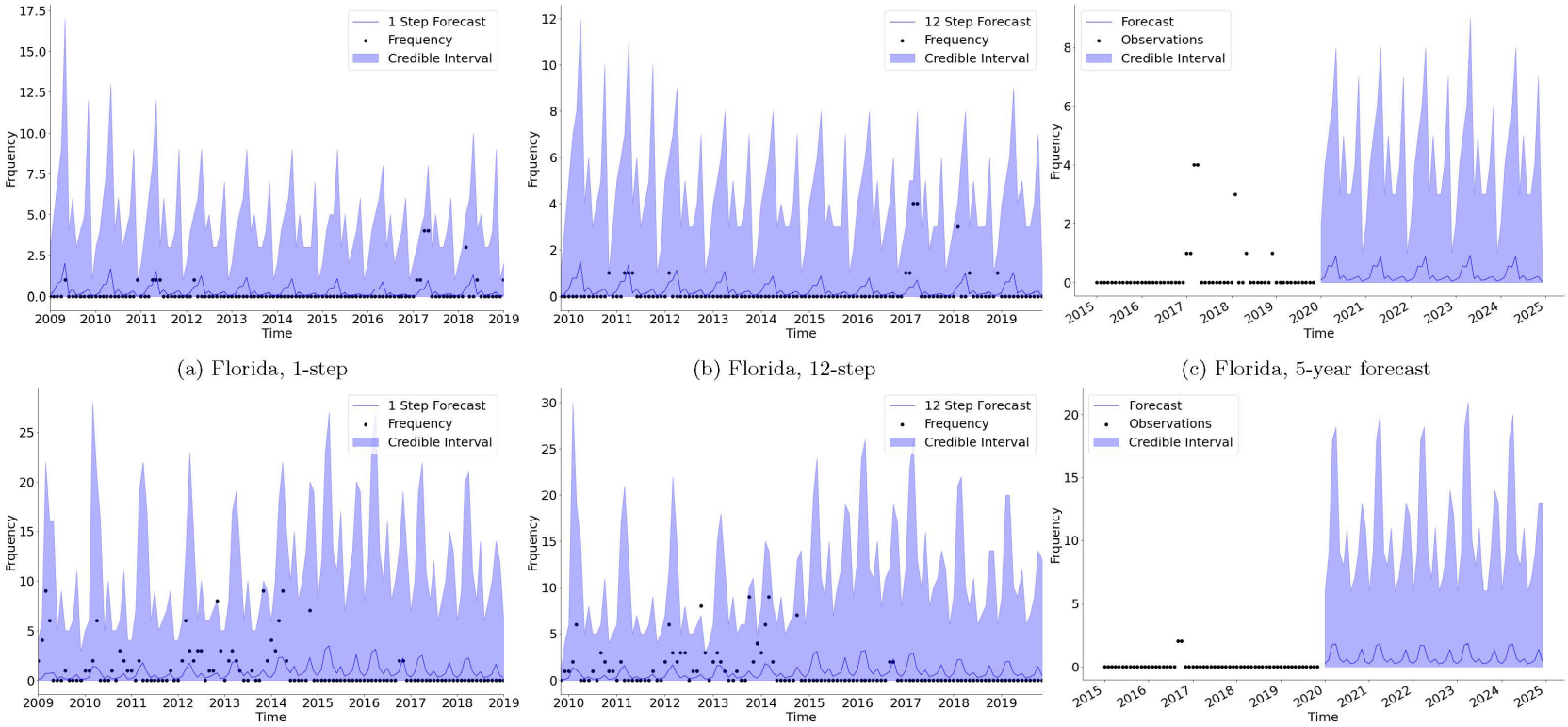


Figure 6. The 1-step (left) and 12-step (mid) mean out-of-sample forecast and the 99% credible intervals, and the 5-year forecasts into the future (right) for Florida (upper panel) and Georgia (lower panel).



(a) Florida, 1-step (b) Florida, 12-step (c) Florida, 5-year forecast

(d) Georgia, 1-step (e) Georgia, 12-step (f) Georgia, 5-year forecast

After examining the figures of the representative states, we now turn to a more comprehensive evaluation of the predictive accuracy of the model. First, we look at the Mean Absolute Deviation (MAD). The MAD for state i and forecast horizon k is given by:

$$\text{MAD}_{i,k} = \frac{1}{T_k - t_k + 1} \sum_{s=t_k}^{T_k} \left| n_{i,t} - \hat{n}_{i,t,k} \right|, \tag{5}$$

where t_k and T_k are the first and the last time point of the k-month ahead out-of-sample forecasts, respectively, and $\hat{n}_{i,t,k}$ is the k-month-ahead mean forecast of $n_{i,t}$. For instance, for $k = 1$, we perform the out-of-sample forecast from January 2005 to January 2019. Hence, we have $t_k = 181$ (the 181^{st} observation) and $T_k = 361$. The MADs are calculated for the two regions as well, where we aggregate the wildfire frequencies $n_{i,t}$ over all relevant states for each t, and obtain the fitted frequency using the regional model.

Table 3 shows the MADs for the two regions over the 12 out-of-sample forecast horizons. We see that the Western region has larger MADs. This is not surprising, as the Western region has overall a much higher number of wildfires over the sample. Further, the MADs are slightly increasing over the forecasting horizon. Table 4 and Table 5 display the state-specific MADs for the Western and the Eastern region, respectively. We see that the state-specific MADs are in general very small, which indicates a satisfying forecasting performance of the proposed model. Moreover, similar to the regional level results, the MADs tend to be larger for states with larger number of wildfires.

Table 3. The Mean Absolute Deviation (MAD) for the whole Western and the Eastern regions over the out-of-sample forecasting horizons k = 1, 2, ..., 12 months.

k	Western Region	Eastern Region
1	8.01	2.57
2	8.23	2.59
3	8.38	2.63
4	8.58	2.68
5	8.66	2.67
6	8.74	2.65
7	8.75	2.69
8	8.81	2.69
9	8.62	2.71
10	8.73	2.70
11	8.93	2.75
12	9.14	2.80

4.3 Wildfire loss severity forecasting

In this subsection we fit the DLM loss severity model introduced in Section 3.3 to the observed losses in the SHELDUS data. For each state i and time t, $y_{i,t}$ is equal to 0 if there was no wildfire in that month, or the logarithm of the aggregate reported property damage (adjusted to 2019 price level) divided by the number of wildfire otherwise. Similar to the DCMM model, we only rely on historical wildfire severity and monthly seasonal factors to forecast their future values. Nevertheless, differing from the multi-scale specification on frequency modeling, we fit the severity data of each state or region separately. The joint modeling of wildfire severity (which consist of excessive zeros and low frequency but very large positive numbers) is a challenging task in itself, and is left for future research.

Since the single-scale dynamic linear model is a mature model and has been applied in numerous studies, we briefly illustrate its goodness-of-fit and the forecasting performance with the two regions and California and Texas. The results of other states are qualitatively similar. First, Figure 7 displays the fitted (forward filtered) mean and 99% credible intervals and the 5-year forecast of the average log wildfire severity of the Western and the Eastern region, respectively. Several observations can be made. First, there are a moderate degree of seasonal effect in the severity series, but not as strong as that of the frequency data. Second, the 99% credible intervals are sufficient to cover most of the non-zero damages, but the intervals are not wide enough to include 0. This is not surprising since the dynamic linear model does not have a separate component to account for the excessive zeros.

Further, it assumes a Gaussian distribution of the log severity, and is thus not able to capture the excessive volatility in the data.

Figure 8 displays the in-sample fits and 5-year forecasts for California and Texas, respectively. Overall we see very similar goodness-of-fit and patterns in projections compared to the regional level. Also, the average log damages are higher in California than in Texas. This could be due to differences in landscape, population density, or price level. A model including related regressors would be helpful to better understand such cross-sectional relationships. However, as stated before, such forecasting is difficult when the number of regressors is large. Therefore, we leave this topic for future research.

Table 4. The Mean Absolute Deviation (MAD) of the Western states over the out-of-sample forecasting horizons k = 1, 2, ..., 12 months.

k	AK	AZ	CA	CO	HI	ID	MT	NV
1	0.69	0.39	2.30	1.28	0.20	0.70	0.58	0.41
2	0.69	0.40	2.25	1.29	0.20	0.70	0.60	0.40
3	0.70	0.40	2.30	1.29	0.20	0.70	0.59	0.41
4	0.70	0.40	2.29	1.30	0.20	0.70	0.59	0.41
5	0.72	0.40	2.31	1.30	0.19	0.70	0.59	0.40
6	0.71	0.40	2.34	1.32	0.19	0.70	0.58	0.40
7	0.70	0.41	2.35	1.31	0.20	0.72	0.53	0.40
8	0.70	0.41	2.33	1.32	0.19	0.73	0.55	0.41
9	0.69	0.40	2.30	1.30	0.19	0.73	0.54	0.39
10	0.67	0.40	2.40	1.30	0.19	0.72	0.54	0.40
11	0.68	0.40	2.43	1.30	0.19	0.72	0.54	0.40
12	0.69	0.40	2.38	1.30	0.19	0.71	0.54	0.40
k	NM	OK	OR	TX	UT	WA	WY	
1	1.94	0.78	0.52	3.85	1.24	1.06	0.36	
2	1.94	0.80	0.52	3.86	1.22	1.06	0.36	
3	1.96	0.79	0.52	3.88	1.19	1.05	0.36	
4	1.95	0.80	0.52	3.95	1.23	1.06	0.36	
5	1.93	0.80	0.53	3.96	1.22	1.06	0.36	
6	1.92	0.80	0.54	3.98	1.23	1.07	0.35	
7	1.93	0.78	0.53	3.94	1.21	1.07	0.35	
8	1.92	0.79	0.52	3.97	1.27	1.09	0.36	
9	1.87	0.77	0.53	3.93	1.21	1.06	0.36	
10	1.88	0.76	0.52	3.91	1.22	1.06	0.36	
11	1.89	0.77	0.51	3.93	1.23	1.05	0.35	
12	1.88	0.78	0.50	3.93	1.19	1.04	0.36	

Table 5, The Mean Absolute Deviation (MAD) of the Eastern states over the out-of-sample forecasting horizons

k	AR	FL	GA	KS	LA	MD	MI	MN	MO
1	0.33	0.48	0.98	0.16	0.26	0.04	0.13	0.13	0.31
2	0.33	0.49	0.99	0.16	0.25	0.04	0.13	0.12	0.31
3	0.34	0.49	1.00	0.16	0.27	0.04	0.13	0.13	0.32
4	0.33	0.49	1.01	0.16	0.26	0.04	0.13	0.12	0.31
5	0.32	0.47	1.02	0.16	0.26	0.04	0.13	0.12	0.30
6	0.32	0.44	1.02	0.16	0.25	0.04	0.13	0.12	0.30
7	0.32	0.44	1.02	0.16	0.25	0.04	0.13	0.12	0.30
8	0.32	0.43	1.02	0.16	0.25	0.04	0.13	0.12	0.30
9	0.32	0.43	1.02	0.16	0.25	0.04	0.13	0.12	0.30
10	0.32	0.43	1.02	0.16	0.25	0.04	0.13	0.12	0.30
11	0.32	0.43	1.03	0.16	0.25	0.04	0.13	0.12	0.30
12	0.31	0.43	1.03	0.15	0.25	0.04	0.13	0.12	0.29
k	NE	NJ	NY	NC	SC	VA	WV	WI	
1	0.23	0.14	0.05	0.09	0.08	0.23	0.35	0.48	
2	0.23	0.14	0.05	0.09	0.08	0.23	0.36	0.49	
3	0.24	0.14	0.06	0.09	0.08	0.23	0.34	0.51	
4	0.23	0.14	0.05	0.09	0.08	0.22	0.34	0.49	
5	0.23	0.14	0.05	0.09	0.08	0.22	0.31	0.49	
6	0.23	0.14	0.05	0.09	0.08	0.22	0.29	0.48	
7	0.23	0.14	0.05	0.09	0.08	0.22	0.30	0.47	
8	0.23	0.14	0.05	0.09	0.08	0.22	0.29	0.46	
9	0.23	0.14	0.05	0.09	0.08	0.22	0.28	0.47	
10	0.23	0.14	0.05	0.09	0.08	0.21	0.28	0.46	
11	0.24	0.14	0.05	0.09	0.08	0.21	0.28	0.45	
12	0.23	0.14	0.05	0.09	0.08	0.21	0.29	0.44	

5. Applications in insurance-linked securities pricing

In this section we perform the hedging practice of wildfire risk with catastrophe bonds. We first introduce the wildfire CAT bond payoff structure and the pricing methods. The (random) aggregate loss in state i up to forecast horizon k is given by:

$$L_{i,k} = \sum_{u=1}^{k} n_{i,T+u} \times y_{i,T+u},$$

(6)

where T is the last observation in the sample. The distribution of $L_{i,k}$ can be obtained via simulation using the predictive distributions of claim frequencies and average claim amounts in Equation (6). Typical CAT bonds are usually priced at spreads over LIBOR, and the loss of principal is determined by a pre-specified loss trigger and an exhaustion point (Zhu 2017). For example, consider a CAT bond written on the aggregate wildfire loss in state i over k months. Denote by $p_{i,k}$, $B_{i,k,1}$, and $B_{i,k,2}$ the bond price, loss trigger, and exhaustion point, respectively. If $L_{i,k}$ is less than $B_{i,k,1}$, then the return to investor at maturity time k is given by:

$$R_{i,k} = p_{i,k} + I_{i,k}, \quad \text{with} \quad I_{i,k} = p_{i,k}\left(e^{r} - 1 + \tilde{l}_{i,k}\right),$$

where $I_{i,k}$ is the interest plus spread premium, with r the LIBOR rate[11] and $\tilde{l}_{i,k}$ the spread premium rate. On the other hand, if $L_{i,k}$ is larger than $B_{i,k,1}$, then the time k return to the investor will be given by

$$R_{i,k} = p_{i,k}\left(1 - f\left(L_{i,k}\right)\right) + I_{i,k},$$

with the loss ratio $f\left(L_{i,k}\right)$ given by:

$$f\left(L_{i,k}\right) = \max\left\{0, \min\left[L_{i,k} - B_{i,k,1}, B_{i,k,2} - B_{i,k,1}\right]\right\} / \left(B_{i,k,2} - B_{i,k,1}\right).$$

In other words, the principal payment will be reduced when $L_{i,k} > B_{i,k,1}$. Moreover, when the aggregate loss exceeds the exhaustion point, the investor will not receive any principal at maturity. However, the interest and spread premium will be paid regardless of the realized loss. In the existing literature, there are several pricing methods of CAT bonds. Specifically, let $\mathrm{l}_{\mathrm{i,k}} = E_{\mathbf{P}}\left[f\left(L_{i,k}\right)\right]$ be the expected loss ratio under the physical measure $\mathbf{P}$, the methods to calculate $\tilde{l}_{i,k}$ include:

$$\tilde{\boldsymbol{l}}_{\boldsymbol{i,k}} = \boldsymbol{b_0} + \boldsymbol{b_1 l_{i,k}},$$

(7)

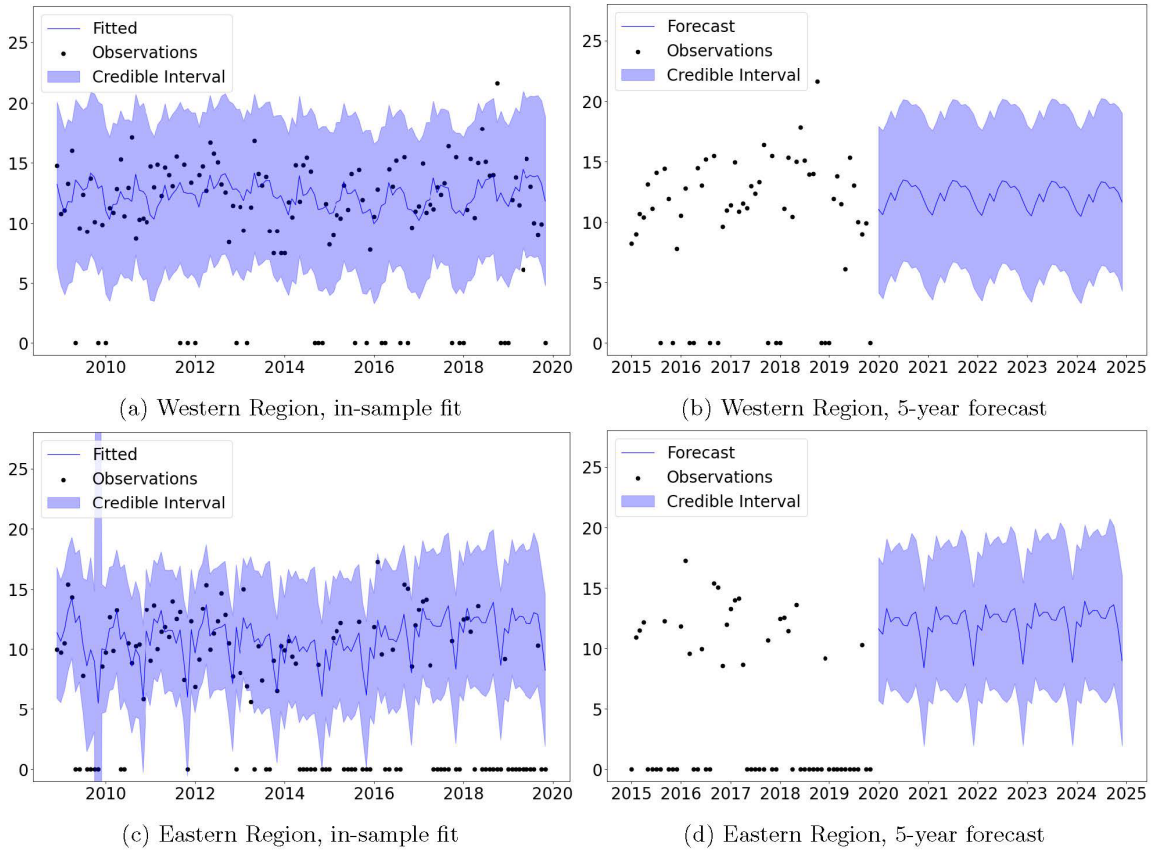
[11] For ease of exposition, we assume that the LIBOR is constant at r for any maturity in this proposal.

$$\tilde{l}_{i,k} = b_0 + b_1 l_{i,k} + b_2 l_{i,k}^2, \tag{8}$$

$$\tilde{l}_{i,k} = l_{i,k} + b_0 \sqrt{l_{i,k}(1 - l_{i,k})}, \tag{9}$$

$$\tilde{l}_{i,k} = b_0 \, l_{i,k}^{b_1}. \tag{10}$$

Figure 7. The fitted mean and the 99% credible intervals (left), and the 5-year forecasts (right) of the average log property damage of wildfire for the Western (upper panel) and Eastern region (lower panel).



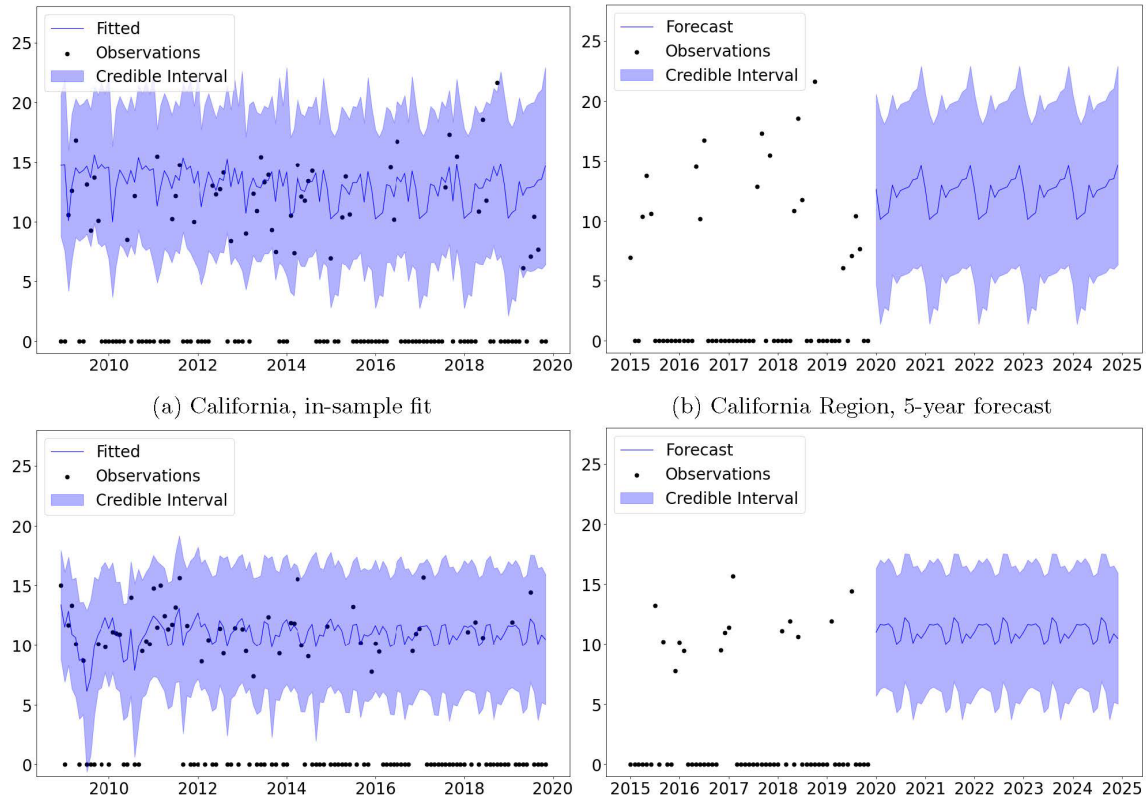
(a) Western Region, in-sample fit

(b) Western Region, 5-year forecast

(c) Eastern Region, in-sample fit

(d) Eastern Region, 5-year forecast

Figure 8. The fitted mean and the 99% credible intervals (left), and the 5-year forecasts (right) of the average log property damage of wildfire for California (upper panel) and Texas (lower panel).



(a) California, in-sample fit

(b) California Region, 5-year forecast

(c) Texas, in-sample fit

(d) Texas, 5-year forecast

Formula (7) is the linear premium principle, which assumes that the spread consists of a fixed level b_0 plus a multiple b_1 of the expected loss. Although the linear premium principle has the simplest form and cannot capture any nonlinear dependency, it is easy to interpret and thus widely accepted by the industry. Formulas (8) to (10) are extensions of the linear premium principle, along different directions. Formula (8) is a quadratic extension to the linear principle, which includes an additional multiple b_2 of the square of the expected loss. Formula (9) is the pricing approach originally developed by the ILS fund Fermat Capital (Trottier, Charest and others 2018). The parameter b_0 can be interpreted as a kind of "insurance-linked security Sharpe ratio." Finally, Formula (10) is introduced by (Major and Kreps 2002). It assumes that the spread is a power function of the expected loss, and thus allows for nonlinear dependencies between the expected loss and the spread.

5.1 Wildfire risk hedging analysis

To calculate the spread premium for given state i and horizon k, we need to compute the predictive distribution of $L_{i,k}$ and the coefficients. As an illustration, we let $k = 12$ (1 year), and focus on three Western states: California, Colorado, and Oregon in this report. Results for other states and regions can be obtained using the same method, and are available upon request. Figure 9 displays the predictive distribution of $L_{i,k}$ for the three states. The distribution of the total aggregate loss of the three states are also shown in Figure 9(d). From the figure, it is clear that the predictive distributions of the 1-year aggregate loss are all heavily skewed to the right. In other words, while the predictive loss is small or moderate in most of the scenarios, there are a number of scenarios in which the predictive loss is huge. In order to gain a better insight of the main part of the distributions, we show the truncate the predictive distributions shown in Figure 9 at the 95% quantile, and display the truncated distributions in Figure 10. Even for the truncated distributions, we observe a very obvious skewness: there is a large probability of having no loss, and non-negligible probabilities of having rather higher losses. We also report their mean, standard deviation, coefficient of variation, skewness, and the 95%, 99% and the 99.5% quantiles in Table 6. The skewness of a sample x_n, $n = 1, \ldots, N$ is defined as $\frac{\sum_n (x_n - \mu)^3}{\sigma^3}$, where μ and σ are the mean and the standard deviation of the sample, respectively. From the statistics, we see that the loss distributions have rather heavy tails. Take California for example, the 95% quantile is more than 3 times larger than the mean forecast, and the 99.5% quantile is 30 times larger. Hence, without any risk mitigation, an insurance company assuming wildfire risk in California may need to pay 30 times more than expected. This number is 28.3 for Colorado and 38.1 for Oregon. Therefore, it is crucially important for insurance companies to have options to transfer their risk exposures, for example, through reinsurance or CAT bonds. Next, we proceed to the determination of the coefficients in the pricing formulas (7) to (10). (Zhu 2017) estimates these coefficients using the CAT bonds data from the quarterly and annual reports published by Lane Financial LLC.[12] We will rely on these estimated coefficients, which are reported in Table 7, for the analysis in this project.

[12] http://www.lanefinancialllc.com/.

Table 6. Summary statistics of the 1-year future aggregate wildfire loss in California, Colorado, Oregon, and the total of the three states.

Statistics (in $billions)	Mean	St.D.	C.V.	Skewness	95% Quantile	99% Quantile	99.5% Quantile
California	12.89	65.55	5.09	14.21	44.15	224.89	391.80
Colorado	0.91	4.28	4.71	13.00	3.46	15.94	25.71
Oregon	0.08	0.50	6.66	14.67	0.23	1.62	3.05
Total	13.87	65.73	4.74	14.14	47.54	226.20	394.26

Figure 9. The predictive distribution of 1-year aggregate wildfire loss in California, Colorado, Oregon, and the total of the three states.

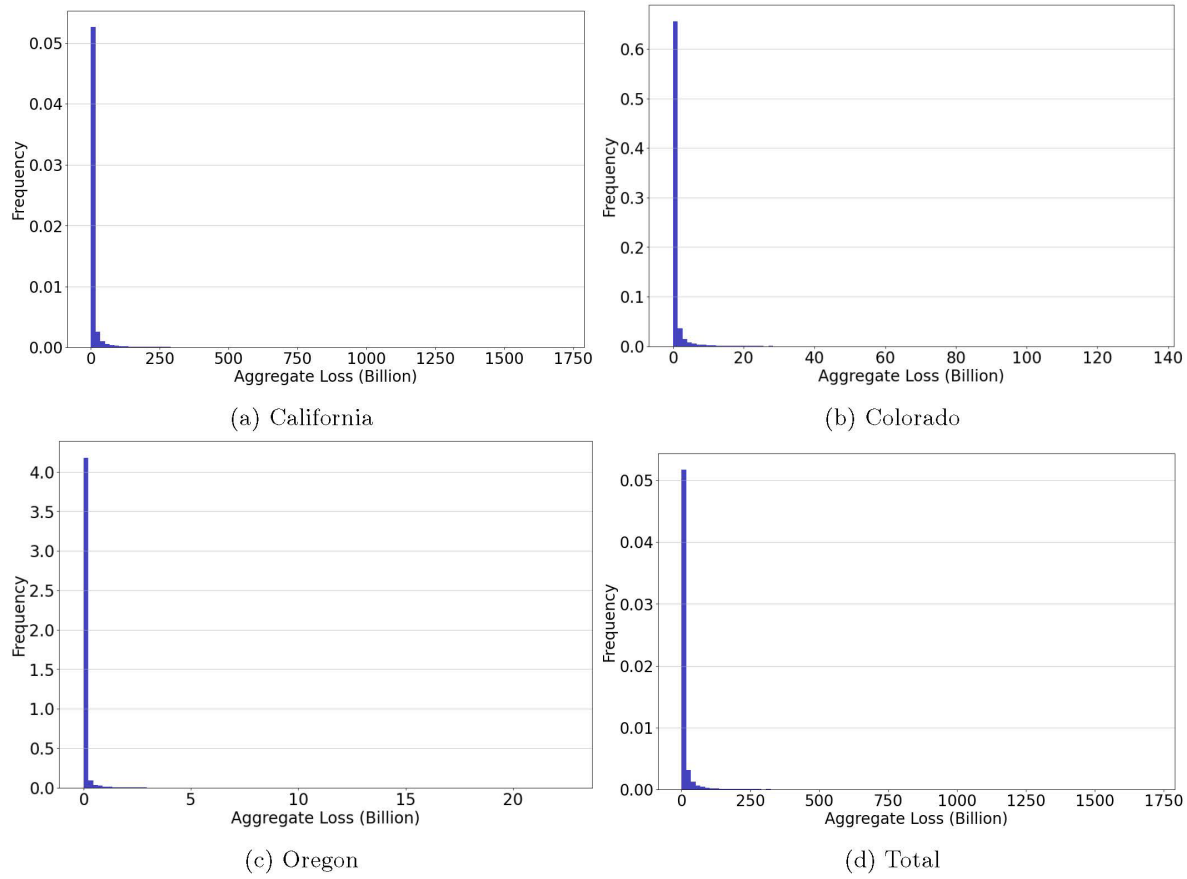


Table 7. The coefficient values of the pricing formulas Equations (7) to (10).

Pricing Formula	(7)	(8)	(9)	(10)
b_0	3.14	3.63	0.43	5.27
b_1	1.98	0.85		0.33
b_2		0.32		

Given the simulated aggregate loss from the dynamic Bayesian models and the estimated coefficients, we calculate the spread premium rates using the expected loss ratio of California, Colorado, Oregon, and the aggregate loss of the three states. In the calculation, we consider two payoff structures of the CAT bond. In the first structure, we let B_1 (the trigger point) and B_2 (the exhaustion point) be the 90% and 99% quantiles of the corresponding loss, respectively. In other words, the principal repayment of each CAT bond will start to decrease when the realized underlying loss is higher than its 90% quantile, and will be reduced to zero when the realized loss exceed its 99% quantile. In the second structure, we let B_1 and B_2 be the 95% and the 99% quantiles, respectively. In this case, only the more extreme risk will be covered.

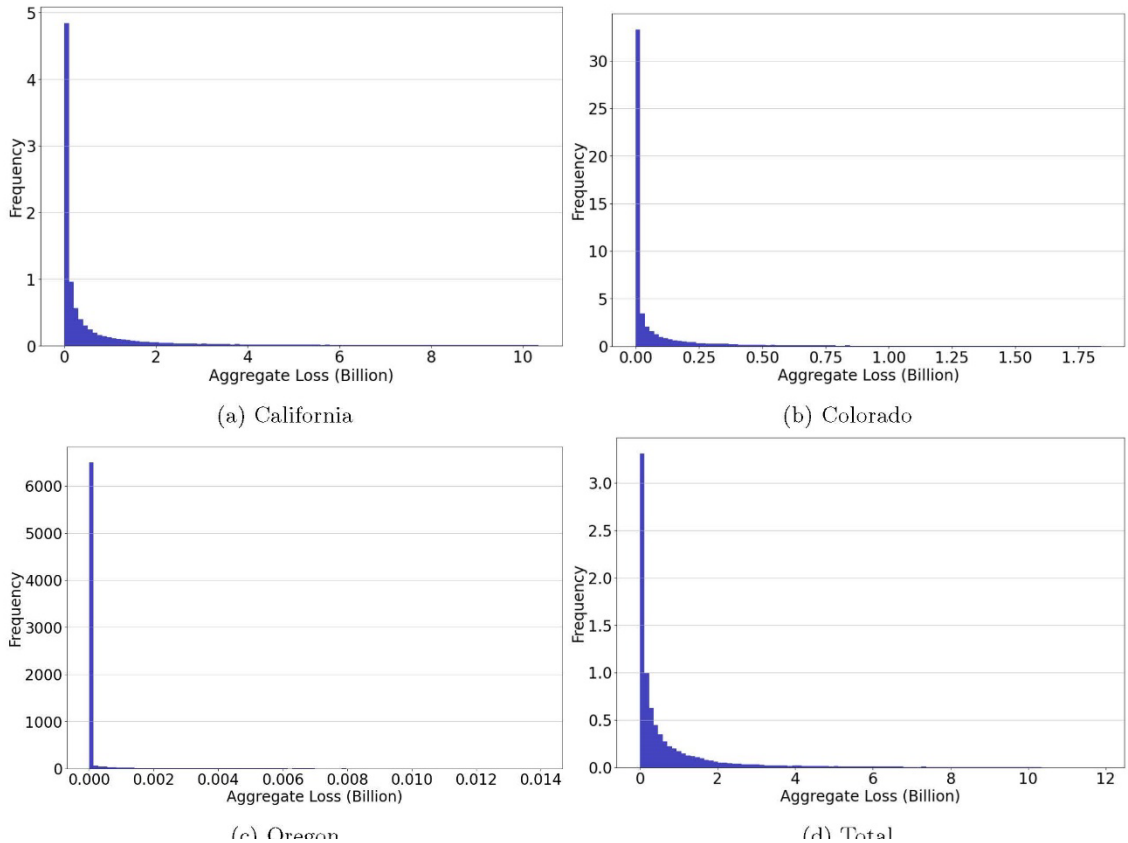
The two structures above resemble a typical CAT bond, as the investors will receive the full principal repayment in most of the scenarios, and will only assume the tail risk. Given this setup, the spread premium rates $\tilde{l}_{i,k}$ are shown in Table 8. First, we see that the premium rate varies substantially with the pricing formulas, ranging from 0.08% to 5.55%. This is not surprising, because the pricing formulas are fundamentally different. For example, while the Formula (7) is linear in the expected loss ratio, Formula (10) relies on the b_1-th moment of the expected loss ratio, and is thus highly nonlinear in the expected loss. Second, the premium rates are almost identical across locations. This is because the setup of the CAT bond is identical for all locations (based on the same quantiles), and thus their expected loss ratios are rather similar: They range from 2.46% to 2.82% for the first payoff structure, and from 2.00% to 2.67% for the second structure.

The spread premium rate $\tilde{l}_{i,k}$ (in terms of %) for a CAT bond written on the 1-year future aggregate wildfire loss in California, Colorado, Oregon, and the total of the three states calculated using the 4 pricing formulas.

Table 8. The spread premium rate $\tilde{l}_{i,k}$ (in terms of %) for a CAT bond written on the 1-year future aggregate wildfire loss in California, Colorado, Oregon, and the total of the three states calculated using the 4 pricing formulas.

(B_1, B_2)	(90%,99%)				(95%,99%)			
Pricing Formula	(7)	(8)	(9)	(10)	(7)	(8)	(9)	(10)
California	3.19	3.65	0.10	5.57	3.18	3.64	0.08	5.54
Colorado	3.20	3.65	0.10	5.58	3.18	3.65	0.08	5.55
Oregon	3.19	3.65	0.09	5.56	3.18	3.64	0.08	5.54
Total	3.19	3.65	0.10	5.57	3.18	3.65	0.08	5.55

Figure 10. The truncated predictive distribution (at the 95% quantile) of 1-year aggregate wildfire loss in California, Colorado, Oregon, and the total of the three states.



Given the spread premium rates, we are finally able to calculate the payoff of the CAT bond, and evaluate their effectiveness in wildfire risk reduction. In this analysis, we assume that there is a representative insurer in each state that assumes all the wildfire risk in that state. In other words, for the representative insurer in California, the expected claim amount over the next year is $11.73 billion, and the 99% quantile is $202.87 billion, etc. For each insurer, we consider CAT bonds with the two payoff structures discussed above. Further, we compare the hedge effectiveness of the CAT bond with two hedging amount: we assume that the insurer can choose to write a CAT bond with principal equal to either the 95% or the 99% quantile of its 1-year future loss. To summarize, we evaluate the hedging effectiveness of the CAT bond under 4 scenarios, where the insurer writes a CAT bond with:

1. Face value equal to the 95% quantile of its 1-year future loss and $(B_1, B_2) = (90\%, 99\%)$ quantiles.

2. Face value equal to the 95% quantile of its 1-year future loss and $(B_1, B_2) = (95\%, 99\%)$ quantiles.

3. Face value equal to the 99% quantile of its 1-year future loss and $(B_1, B_2) = (90\%, 99\%)$ quantiles.

4. Face value equal to the 99% quantile of its 1-year future loss and $(B_1, B_2) = (95\%, 99\%)$ quantiles.

The summary statistics of the simulated 1-year hedged loss for each insurer are shown in Table 9. For illustrative purpose, we only report the results using the linear pricing formula (7). Results using other pricing formulas are qualitatively similar, and are available upon request. Several observations can be drawn from the hedging results. First, the hedged loss in general has a higher mean value than the unhedged loss shown in Table 6, due to the premium that the insurer has to pay to the investors. Second, and more importantly, the standard deviation and the 99% and the 99.5% quantiles of the hedged loss are much smaller than the unhedged loss, indicating much lower volatility and tail risk for the hedged loss. This is also the case for the 95% quantile for all scenarios except scenario 2. Therefore, if the insurer is regulated to set the 99.5% value-at-risk of its future loss over the one year horizon as the minimum capital, then hedging with the CAT bonds could substantially reduce its capital requirement. This is because in Scenario 2, the principal payments just starts to reduce when the loss exceeds its predicted 95% quantile, and thus the insurer does not receive sufficient compensation at the 95% to offset the premium it pays. Third, the skewness of the hedged loss is in general higher than that of the unhedged loss for all states. This indicates that the hedged loss still has a heavy tail. This is not surprising because the CAT bonds only mitigate the tail risk, and did not change the main part of the loss distribution. Also, the magnitude of the skewness is larger because the the hedged loss has a higher expectation and a lower standard deviation than the unhedged loss. Finally, the coefficient of variation of the hedged losses are smaller than the unhedged ones in all cases. To summarize, while writing the CAT bond moderately increases the mean loss (and thus the liability) of the insurer, it can substantially reduce the volatility and the tail risk.

Table 9. Summary statistics of the hedged loss over the 1-year horizon in California, Colorado, and Oregon under the 4 scenarios

State	Scenario	Mean	St.D.	C.V.	Skewness	95% Quantile	99% Quantile	99.5% Quantile
California	1	13.06	62.82	4.81	16.10	39.87	173.51	345.28
	2	13.32	62.99	4.73	15.96	45.40	173.51	345.28
	3	13.90	48.68	3.50	21.69	24.76	25.99	179.73
	4	15.19	49.15	3.24	21.00	38.81	50.50	179.71
Colorado	1	0.92	3.92	4.24	14.33	3.10	12.61	22.38
	2	0.95	3.94	4.16	14.13	3.57	12.61	22.38
	3	0.98	2.84	2.90	20.77	1.92	2.02	10.36
	4	1.08	2.89	2.68	19.59	2.97	3.92	10.36
Oregon	1	0.08	0.48	6.24	15.41	0.21	1.40	2.83
	2	0.08	0.48	6.16	15.35	0.24	1.40	2.83
	3	0.09	0.37	4.23	20.75	0.12	0.12	1.49
	4	0.09	0.37	3.90	20.40	0.21	0.28	1.49

5.2 Basis risk analysis

In Section 5.1, we evaluated the hedging effectiveness using indemnity CAT bonds. In other words, the CAT bonds are written on the risk specific to the insurer, and their payoffs are perfectly linked to the realized loss of the insurer. In this subsection, we relax this assumption and examine the hedging effectiveness of *index-based* CAT bonds. Specifically, we now assume that the trigger point and exhaustion point of all CAT bonds are based on the predictive distribution of the total loss of the three states combined. In this case, the CAT bond payments are no longer perfectly related to the realized loss of each insurer. Instead, they are related to an "index," which is assumed to be the total loss of the three states in this case. When hedging with such index-based bonds, the insurers are bearing *basis risk*, i.e., the risk of mismatch between the CAT bond payments and their realized loss.

Similar to Section 5.1, we consider 4 scenarios with different face values and (B_1, B_2). Note that, although now (B_1, B_2) are related to the total loss, the face value is still assumed to be based on their own liability. This resembles the situation in reality, where insurers are free to determine how much CAT bond they want to purchase/issue. The summary statistics of the hedged loss using the index-based bonds are shown in Table 10. Compared to the benchmark case in Table 9, all numbers are larger, indicating a worse hedging result. The magnitude of increase ranges from 10% to 20%, which seems to be reasonable, especially for Colorado and Oregon, where the risk exposures are low. For these states, the frequency of wildfire is much more volatile and has more zeros than California, and thus more difficult to predict. As a result, the CAT bond written solely on the loss of one of these states may be less liquid and more expensive, because they are riskier and only relevant to agents exposed to wildfire risk specific in this state (which may not be many). Therefore, a 10% to 20% margin in exchange of issuing an index-based CAT bond, which is more liquid and less expensive, might be a reasonable choice of many insurers.

Table 10. Summary statistics of the hedged loss over the 1-year horizon in California, Colorado, and Oregon with CAT bonds in the presence of basis risk.

State	Scenario	Mean	St.D.	C.V.	Skewness	95% Quantile	99% Quantile	99.5% Quantile
California	1	13.06	67.36	5.16	14.72	44.54	217.26	387.84
	2	13.32	67.33	5.05	14.73	44.72	217.25	388.74
	3	13.92	72.57	5.21	11.40	48.51	220.23	387.91
	4	15.21	72.05	4.74	11.67	49.34	220.45	389.50
Colorado	1	0.93	4.26	4.60	12.68	3.50	15.93	25.67
	2	0.95	4.28	4.51	12.70	3.53	15.95	25.79
	3	0.99	4.56	4.61	8.44	3.80	16.00	25.47
	4	1.08	4.60	4.24	8.83	3.85	16.16	26.05
Oregon	1	0.08	0.50	6.57	14.59	0.23	1.62	3.05
	2	0.08	0.50	6.46	14.59	0.23	1.62	3.06
	3	0.08	0.54	6.51	11.21	0.27	1.64	3.07
	4	0.09	0.54	5.80	11.47	0.27	1.65	3.09

5.3 Single-Scale model

Up to now, the prediction of wildfire frequencies are based on the multi-scale DCMM model (4), in which spatial dependence is accounted for via the common factors $\boldsymbol{\phi}_t$. In this subsection, we evaluate the importance of spatial dependence by repeating the analysis using state-specific, independent DCMM models (Model (1) without the common factor). The resulting 1- to 12-month MADs are collected in Appendix 9. In general, the MADs are larger with independent DCMM models, which indicates that incorporating the spatial dependence helps increase the predictive accuracy.

The summary statistics of the unhedged and hedged 1-year loss are reported in Table 11 and Table 12, respectively. Compared to the benchmark results in Section 5.1, we see that the predictive distributions of the 1-year loss, both hedged and unhedged, have higher means, and are more volatile in all states. Hence, incorporating the spatial dependence has a substantial impact on the prediction of future loss and the hedging results. Based on the predictive accuracy measure, it seems that the results with spatial dependence are more credible.

Table 11. Summary statistics of the 1-year future aggregate wildfire loss in California, Colorado, Oregon, and the total of the three states using single-scale models.

State	Mean	St.D.	C.V.	Skewness	95% Quantile	99% Quantile	99.5% Quantile
California	13.64	68.37	5.01	13.93	48.13	236.27	413.23
Colorado	1.19	5.70	4.81	13.13	4.43	20.17	34.63
Oregon	0.14	1.07	7.66	17.21	0.37	2.86	5.46
Total	14.97	68.60	4.58	13.78	52.83	238.33	413.77

Table 12. Summary statistics of the hedged loss over the 1-year horizon in California, Colorado, and Oregon with CAT bonds of different characteristics using single-scale models.

State	Scenario	Mean	St.D.	C.V.	Skewness	95% Quantile	99% Quantile	99.5% Quantile
California	Bond 1	13.89	63.54	4.58	15.12	43.64	189.92	366.88
	Bond 2	14.17	63.74	4.50	14.97	49.65	189.91	366.87
	Bond 3	14.83	47.69	3.22	20.91	27.22	28.59	185.67
	Bond 4	16.23	48.28	2.97	20.09	42.22	55.16	185.64
Colorado	Bond 1	1.20	5.24	4.36	14.40	3.97	15.91	30.36
	Bond 2	1.23	5.26	4.27	14.22	4.57	15.91	30.36
	Bond 3	1.27	3.88	3.06	20.16	2.45	2.58	15.20
	Bond 4	1.39	3.94	2.83	19.15	3.82	5.02	15.19
Oregon	Bond 1	0.14	1.04	7.28	17.85	0.34	2.51	5.11
	Bond 2	0.14	1.04	7.20	17.81	0.38	2.51	5.11
	Bond 3	0.16	0.85	5.25	22.38	0.20	0.21	2.71
	Bond 4	0.17	0.85	4.89	22.20	0.35	0.46	2.71

6. Reinsurance analysis

Reinsurance is a traditional approach to transfer climate risk for insurers, and it is currently a more widely used. Here, we will extend the applications of the fitted models described in Section 4 to evaluate the hedge effectiveness of stop-loss reinsurance products. In

particular, we will calculate the **actuarially fair** premium for a stop-loss reinsurance with different thresholds, as well as the summary statistics of the reinsured liabilities. Stop-loss reinsurance caps the loss of an insurer at a fixed level (Kaas, et al. 2008). Formally, for state i, a stop-loss reinsurance with threshold s pays $P_i = L_i - s$ to the insurer if the total insurance loss over a given time period (typically, 1 year) L_i is larger than s, and $P_i = 0$ otherwise. The actuarially fair premium of the reinsurance is given by:

$$p_i = e^{-r} E[P_i], \tag{11}$$

i.e., the expected present value of the reinsurance payment. The end-of-year net reinsured loss (after paying the premium) of the insurer is thus:

$$\tilde{L}_i = \begin{cases} L_i + p_i e^r, & \text{if } L_i < s, \\ s + p_i e^r, & \text{if } L_i \geq s. \end{cases} \tag{12}$$

In this analysis, we will discuss two scenarios of reinsurance with different stop-loss thresholds for each state:

Stop-loss threshold equals 95% of its 1-year predicted loss.

Stop-loss threshold equal 99% of its 1-year predicted loss.

For simplicity, we leave out the subscripts i and k from the symbols. The actuarially fair premiums of the reinsurance, as well as the summary statistics of the net reinsured losses of three states are shown in Table 13. We see that for each state the reinsurance premium with a lower stop-loss level is more expensive, which is intuitive. Moreover, the means of the reinsured losses are the same in both scenarios. This is because the actuarially fair premium is used. Specifically, the expected end-of-year net reinsured loss is given by:

$$\begin{aligned} E[\tilde{L}_i] &= E[L_i - P_i + p_i e^r] \\ &= E\big[L_i - P_i + E[P_i]\big] \\ &= E[L_i] - E[P_i] + E[P_i] \\ &= E[L_i]. \end{aligned}$$

In other words, the expected reinsured loss is identical to the expected unreinsured loss if the reinsurance has an actuarially fair premium. The expected reinsured and unreinsured losses will be different if another premium principle is used. Further, comparing between the two scenarios, we see that the reinsured losses are less volatile and have lower 95% and 99% quantiles under Scenario 1 in which the stop-loss threshold is lower. Finally, compared to the hedged loss using CAT bonds, the reinsured losses in general are less volatile, have a smaller coefficient of variation, but higher 95% and 99% quantiles. Therefore, there is no universal conclusion that can be made about which risk mitigating

strategy is more appealing than the other, and the choice between two should depend on the availability of the products and an insurer's choice of the selection criteria.

Table 13. Summary statistics of the hedged loss over the 1-year horizon in California, Colorado, and Oregon with stop-loss reinsurance.

State	Scenario	Reinsurance Price	Mean	St.D.	C.V.	Skewness	95% Quantile	99.5% Quantile
California	1	6.76	12.84	11.11	0.87	2.50	50.76	50. 76
	2	3.17	12.84	29.02	2.26	5.37	47.17	219.50
Colorado	1	0.45	0.91	0.88	0.97	2.50	3.91	3.91
	2	0.19	0.91	2.18	2.39	5.21	3.64	16.12
Oregon	1	0.05	0.08	0.57	0.75	2.88	0.28	0.28
	2	0.02	0.08	0.21	2.74	6.04	0.25	1.64

7. Case Study: California county-level data

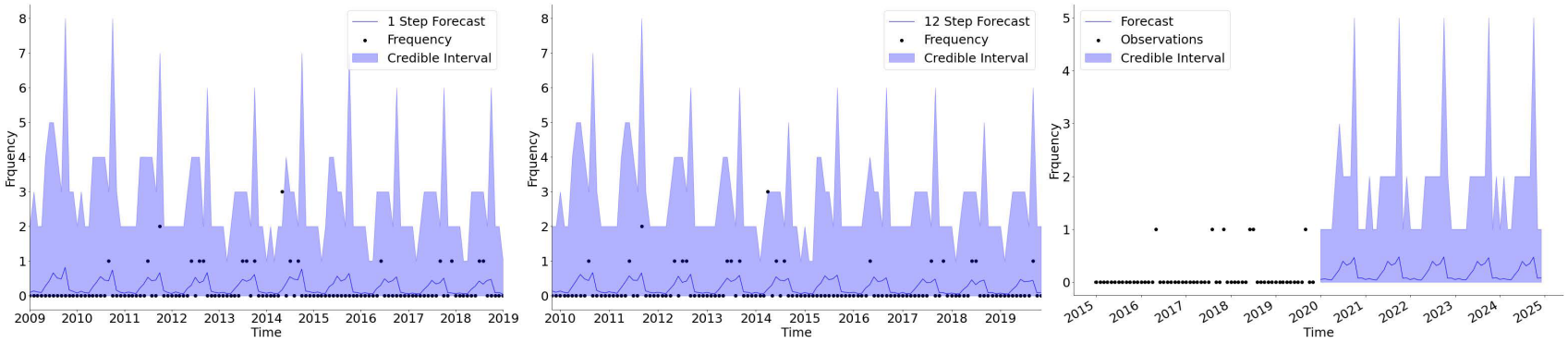
In this section, we apply the analysis in Sections 4 and 5 to the county-level data of California. Similarly, monthly data from January 1989 are used, and counties with 15 historical wildfires or fewer are excluded. After cleaning the data, we are left with 11 counties, which are summarized in Table 14.

Table 14. Counties in California with the numbers in the brackets are the number of wildfires that occurred during January 1989 to December 2019.

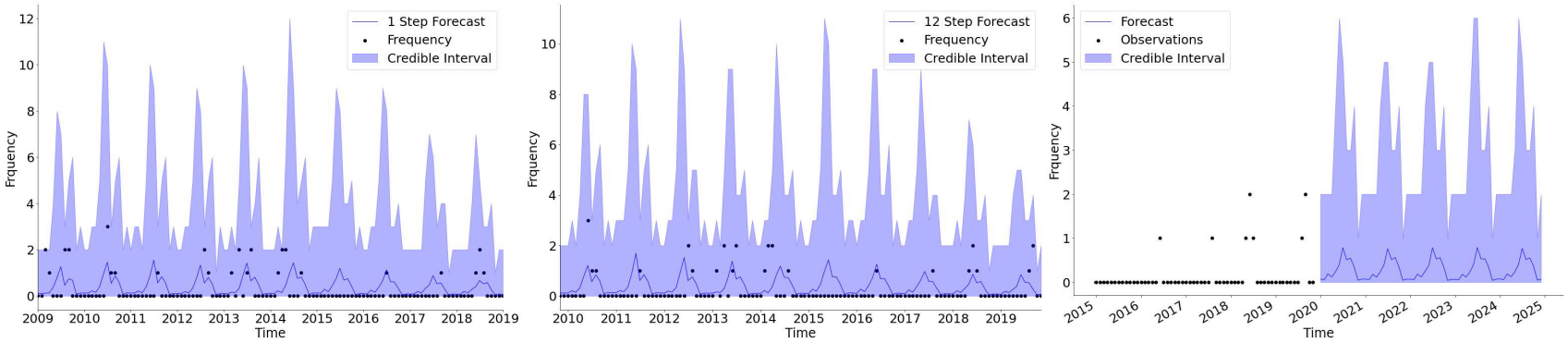
County			
El Dorado (16)	Kern (22)	Los Angeles (21)	Mariposa (17)
Orange (23)	Riverside (99)	San Bernardino (105)	San Diego (77)
Shasta (17)	Tulare (15)	Ventura (17)	

The 1- and the 12-month out-of-sample forecasts for the three more wildfire counties: San Diego, Riverside, and San Bernardino, are shown in Figure 11 and Figure 12. We see that, compared to the state-level data, the county-level frequencies are much lower, with the most non-zero number of claims being one or two. More importantly, similar to the state-level analysis, the 95% credible intervals produced by the DCMM are able to include most of the realized historical frequencies for the three counties shown. The mean absolute deviance of the counties in California over the 1- to 12-month out-of-sample forecast horizons are shown in Table 15. The in-sample fit and the 5-year forecast of the average log property damage of wildfire for the three counties are displayed in Figure 13. We see that the average log damages are concentrated between 8 and 15, or equivalently between \$3,000 and \$3.27 million losses for the three counties reported. Furthermore, all historical observations are within the 95% credible intervals generated by the proposed DCMM.

Figure 11. The 1-step (left) and 12-step (mid) mean out-of-sample forecast and the 99% credible intervals, and the 5-year forecasts into the future

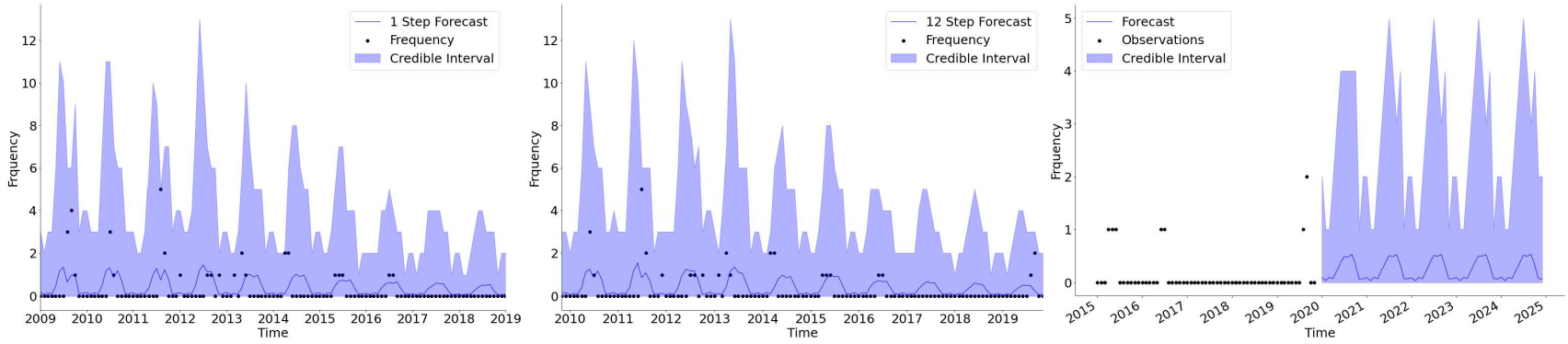


(a) San Diego, 1-step (b) San Diego, 12-step (c) San Diego, 5-year forecast



(d) Riverside, 1-step (e) Riverside, 12-step (f) Riverside, 5-year forecast

Figure 12. The 1-step (left) and 12-step (mid) mean out-of-sample forecast and the 99% credible intervals, and the 5-year forecasts into the future (right) for San Bernardino.



(a) San Bernardino, 1-step

(b) San Bernardino, 12-step

(c) San Bernardino, 5-year forecast

Table 15. The Mean Absolute Deviation (MAD) of the counties in California over the out-of-sample forecasting horizons k=1,2,...,12 months.

k	El Dorado	Kern	Los Angeles	Mariposa	Orange	Riverside
1	0.16	0.18	0.37	0.22	0.20	0.59
2	0.16	0.18	0.37	0.22	0.20	0.58
3	0.16	0.18	0.37	0.22	0.19	0.57
4	0.16	0.18	0.37	0.22	0.19	0.57
5	0.16	0.19	0.37	0.22	0.19	0.57
6	0.17	0.18	0.37	0.22	0.19	0.57
7	0.17	0.19	0.38	0.22	0.19	0.57
8	0.17	0.18	0.37	0.22	0.19	0.57
9	0.17	0.18	0.35	0.21	0.19	0.57
10	0.18	0.18	0.38	0.21	0.19	0.57
11	0.17	0.18	0.38	0.21	0.19	0.57
12	0.17	0.18	0.38	0.21	0.19	0.57
k	San Bernardino	San Diego	Shasta	Tulare	Ventura	
1	0.64	0.43	0.17	0.19	0.18	
2	0.64	0.43	0.17	0.19	0.18	
3	0.65	0.42	0.17	0.20	0.18	
4	0.64	0.42	0.17	0.20	0.18	
5	0.64	0.42	0.17	0.20	0.19	
6	0.64	0.42	0.17	0.20	0.18	
7	0.64	0.42	0.17	0.20	0.19	
8	0.65	0.42	0.17	0.19	0.19	
9	0.65	0.42	0.17	0.18	0.18	
10	0.65	0.41	0.16	0.17	0.18	
11	0.64	0.41	0.16	0.17	0.18	
12	0.65	0.41	0.16	0.18	0.18	

Figure 14 and Figure 15 display the original and the truncated (at 95% quantile) predictive distribution of the 1-year aggregate loss for each of the three counties. The corresponding summary statistics are shown in Table 16. We see that the loss distributions are rather different across counties. In particular, all summary statistics reported (mean, standard, and 95% and 99% quantiles) for Riverside are lower than 10% of those of San Diego. This is because the Cedar fire,[13] which occurred in October 2003 in San Diego and has caused $1.5 billion property damage (adjusted to 2019 price level), boosted the predicted future loss of San Diego.

Next, we perform the hedging analysis described in Sections 5.1 and 5.2 on the county-level loss data. In particular, the four scenarios in Section 5.1 are considered, in which the insurer writes an indemnity CAT bond with different face values and triggers. The hedged results of the three counties are shown in Table 17. Similar to the state-level analysis, we see that hedging with the indemnity CAT bond will lead to higher mean loss because of the premium paid by the insurer. However, the volatility of losses is substantially

[13] Source: https://en.wikipedia.org/wiki/Cedar_Fire.

Figure 13. The fitted mean and the 99% credible intervals (left), and the 5-year forecasts (right) of the average log property damage of wildfire for San Diego (upper panel), Riverside (middle panel), and San Bernardino (lower panel).

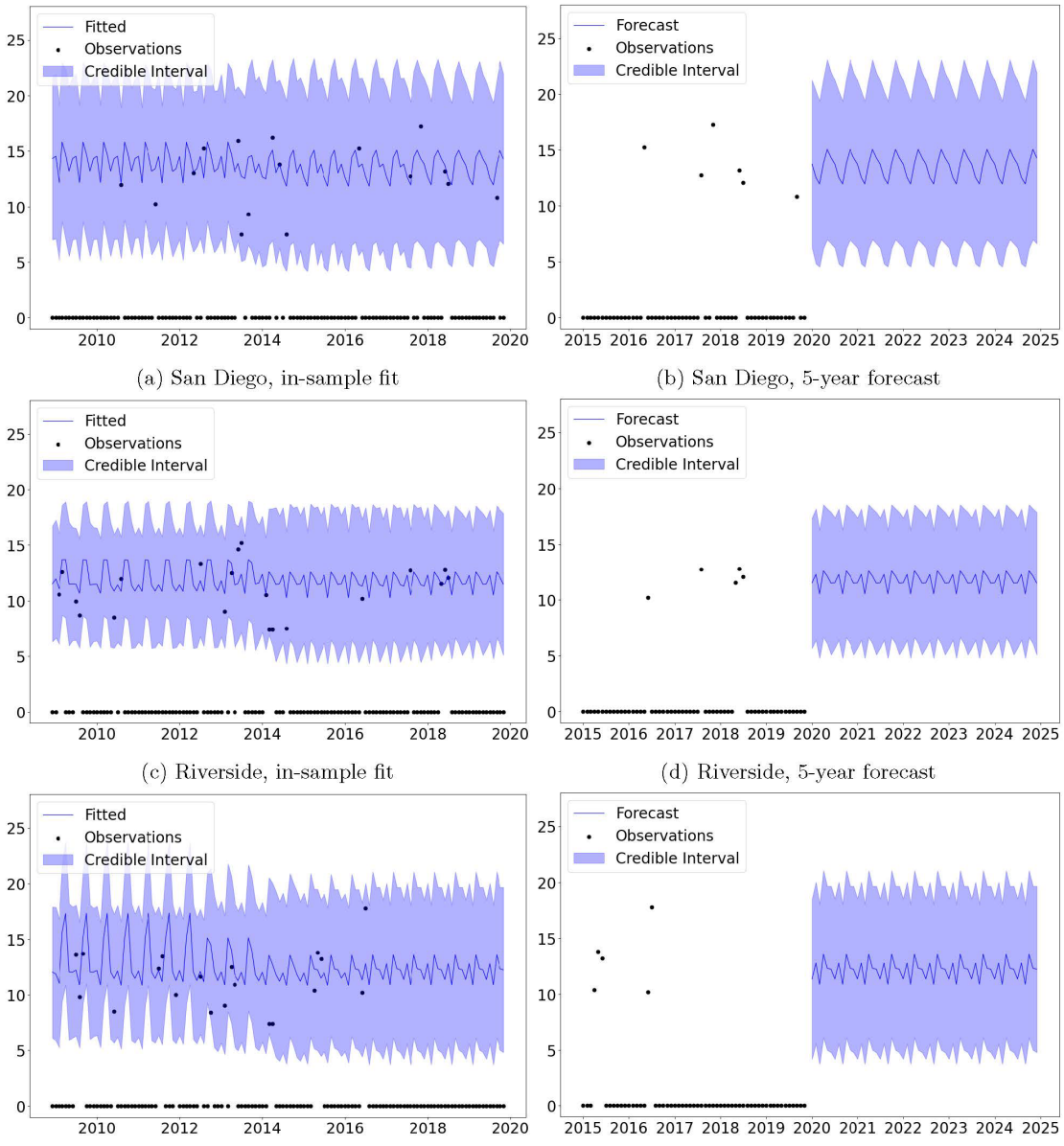


Figure 14. The predictive distribution of 1-year aggregate wildfire loss in San Diego, Riverside, San Bernardino, and the total of the three states.

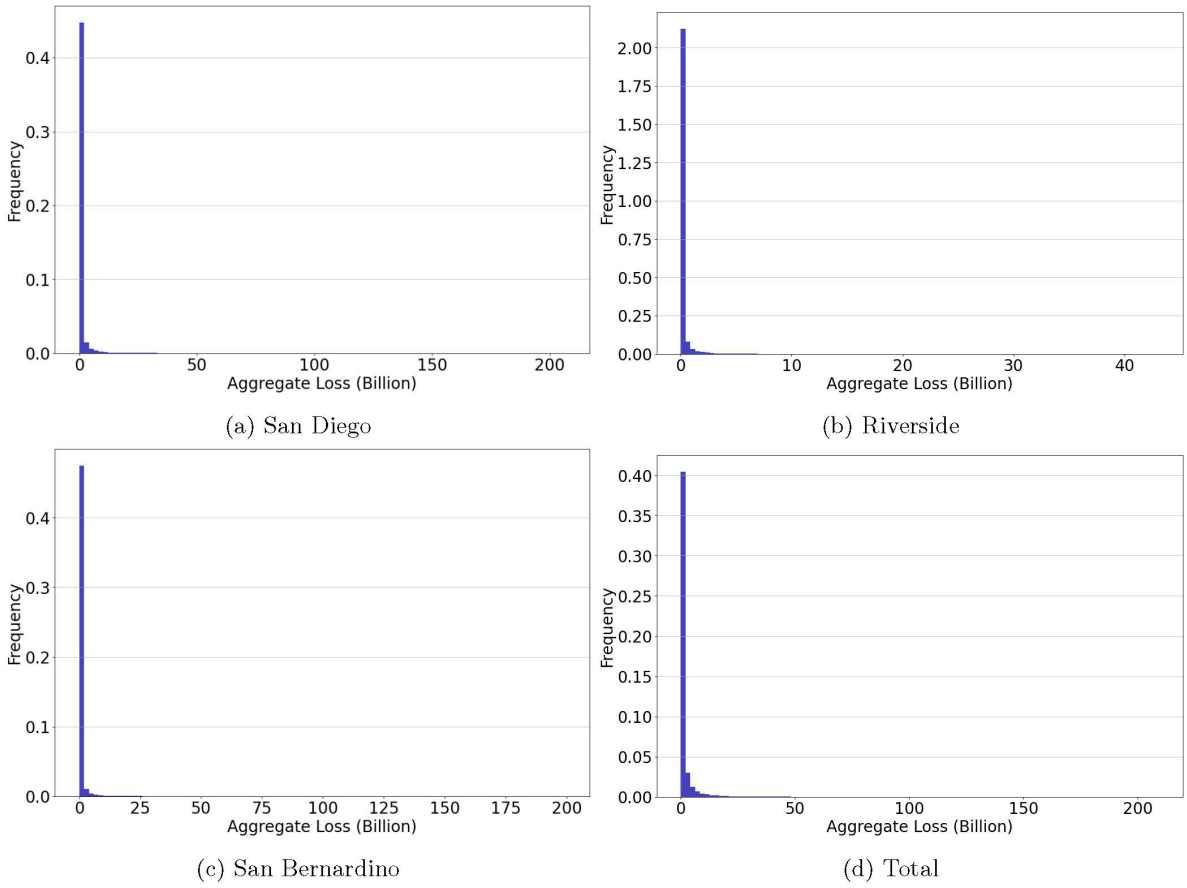
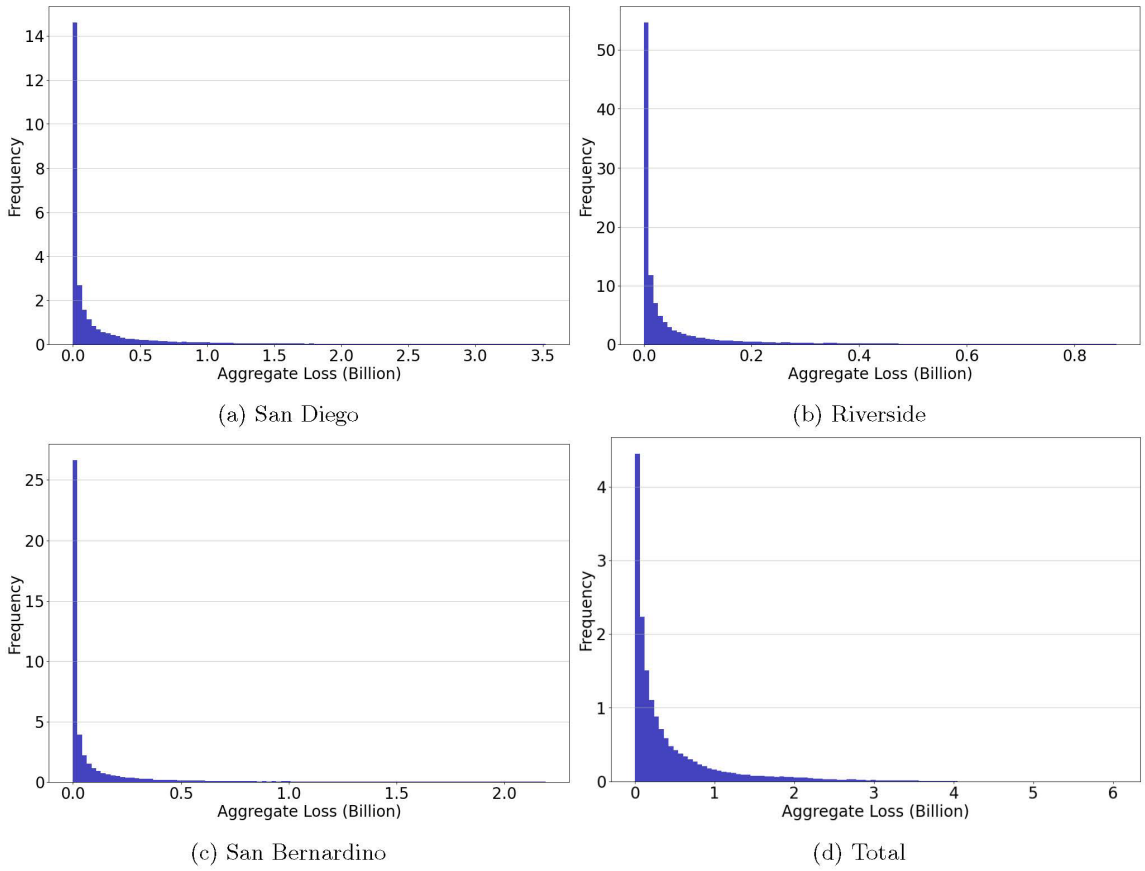


Figure 15. The truncated (at 95% quantile) predictive distribution of 1-year aggregate wildfire loss in San Diego, Riverside, San Bernardino, and the total of the three states.



reduced. Namely, we see smaller standard deviations, and much smaller 99% quantiles, especially in Scenarios 3 and 4 where the face value of the CAT bond is equal to 99% of the projected 1-year loss in each county. Next, we take the basis risk into account, and consider index-based CAT bonds written on the total loss of the three counties combined. The hedged results in the presence of basis risk are shown in Table 18. Again, similar to the state-level analysis, we see that the presence of basis risk increases the standard deviations and the quantiles of hedged losses, and thus leads to a worse hedging results. This is especially the case for Scenarios 3 and 4, when the face value of the CAT bond is higher. In particular, the 99% loss quantiles for Riverside and San Bernardino have increased by 15 times for Scenario 3. Therefore, our analysis indicates that while index-based CAT bonds can have a reasonable mean hedging performance (in terms of mean and standard deviation), the inherent basis risk leads to much worse hedging results for the tail risk and for counties with lower aggregate losses.

Finally, we repeat the analysis in Section 5.3, and compare the unhedged and hedged loss predicted by the single-scale models with the previous predictions using the multi-scale model. The unhedged summary statistics of each county predicted by the single-scale models are shown in Table 19, and the hedged results are shown in Table 20. Again, the

results are qualitatively similar to the state-level analysis. Specifically, we see that the single-scale models predict (much) higher losses, whether hedged or unhedged. This is especially the case for the loss quantiles of Riverside and San Bernardino, which have lower historical losses. Hence, incorporating the spatial dependence has a substantial benefit on predicting future insurance losses and the hedging analysis for county-level data as well.

Summary statistics of the 1-year future aggregate wildfire loss in San Diego, Riverside, San Bernardino, and the total of the three counties.

Table 16. Summary statistics of the 1-year future aggregate wildfire loss in San Diego, Riverside, San Bernardino, and the total of the three counties.

State	Mean	St.D.	C.V.	Skewness	95% Quantile	99% Quantile	99.5% Quantile
San Diego	1.08	6.70	6.22	15.89	3.44	20.66	38.46
Riverside	0.26	1.58	6.01	14.78	0.86	4.82	9.05
San Bernardino	0.83	6.20	7.50	16.95	2.13	16.31	32.07
Total	2.17	9.24	4.27	11.12	8.22	37.70	63.18

Table 17. Summary statistics of the hedged loss over the 1-year horizon in San Diego, Riverside, and San Bernardino under the 4 scenarios.

State	Scenario	Mean	St.D.	C.V.	Skewness	95% Quantile	99% Quantile	99.5% Quantile
San Diego	1	1.10	6.38	5.81	16.87	3.17	17.35	35.15
	2	1.12	6.39	5.72	16.79	3.54	17.35	35.15
	3	1.21	5.00	4.14	22.54	1.92	1.99	18.56
	4	1.31	5.02	3.83	22.10	3.12	4.07	18.56
Riverside	1	0.27	1.49	5.59	15.69	0.79	3.99	8.23
	2	0.27	1.50	5.50	15.60	0.89	3.99	8.23
	3	0.29	1.17	4.03	20.42	0.47	0.49	4.41
	4	0.32	1.18	3.71	19.96	0.77	1.01	4.41
San Bernardino	1	0.84	6.01	7.13	17.57	2.01	14.26	30.01
	2	0.85	6.01	7.06	17.54	2.20	14.26	30.01
	3	0.95	4.92	5.17	21.90	1.22	1.25	16.36
	4	1.02	4.93	4.82	21.73	2.05	2.64	16.36

Table 18. Summary statistics of the hedged loss over the 1-year horizon in San Diego, Riverside, and San Bernardino with CAT bonds in the presence of basis risk.

State	Scenario	Mean	St.D.	C.V.	Skewness	95% Quantile	99% Quantile	99.5% Quantile
San Diego	1	1.09	6.44	5.90	16.44	3.48	19.04	35.15
	2	1.11	6.45	5.79	16.34	3.52	19.31	35.15
	3	1.16	5.49	4.71	16.71	3.80	10.77	18.56
	4	1.30	5.52	4.26	16.41	3.98	12.35	18.56
Riverside	1	0.27	1.55	5.83	14.59	0.87	4.78	8.91
	2	0.27	1.55	5.73	14.54	0.88	4.83	9.03
	3	0.28	1.53	5.41	10.72	0.97	4.69	8.32
	4	0.31	1.54	4.93	10.79	0.99	4.88	8.94
San Bernardino	1	0.84	6.05	7.24	17.25	2.17	15.53	30.34
	2	0.85	6.06	7.13	17.20	2.18	15.74	30.39
	3	0.89	5.33	5.95	17.10	2.50	10.55	18.76
	4	1.00	5.34	5.35	16.96	2.58	12.10	19.20

Table 19. Summary statistics of the unhedged loss over the 1-year horizon in San Diego, Riverside, and San Bernardino.

State	Mean	St.D.	C.V.	Skewness	95% Quantile	99% Quantile	99.5% Quantile
San Diego	1.50	9.80	6.53	16.18	4.63	27.76	54.33
Riverside	0.32	2.19	6.74	17.68	0.98	5.99	11.38
San Bernardino	1.14	9.85	8.63	19.69	2.55	21.45	46.34
Total	2.97	14.10	4.75	12.29	10.73	52.45	90.75

Table 20. Summary statistics of the hedged loss over the 1-year horizon in San Diego, Riverside, and San Bernardino under the 4 scenarios using single-scale models.

State	Scenario	Mean	St.D.	C.V.	Skewness	95% Quantile	99% Quantile	99.5% Quantile
San Diego	1	1.53	9.37	6.13	17.06	4.27	23.31	49.87
	2	1.55	9.39	6.05	16.99	4.77	23.30	49.87
	3	1.67	7.55	4.53	21.82	2.54	2.64	27.59
	4	1.81	7.58	4.18	21.48	4.17	5.48	27.59
Riverside	1	0.33	2.10	6.35	18.61	0.90	5.05	10.44
	2	0.34	2.10	6.26	18.55	1.01	5.05	10.44
	3	0.36	1.72	4.72	23.77	0.55	0.57	5.61
	4	0.39	1.73	4.38	23.45	0.90	1.16	5.61
San Bernardino	1	1.16	9.65	8.29	20.22	2.42	18.99	43.88
	2	1.17	9.65	8.23	20.20	2.63	18.99	43.88
	3	1.32	8.30	6.28	24.12	1.49	1.52	25.68
	4	1.41	8.31	5.90	24.03	2.50	3.22	25.68

8. Conclusion

In this project, we apply an innovative Bayesian multi-scale Dynamic Count Mixture Model (DCMM) to predict location-specific wildfire frequencies. The proposed model is capable of handling the intricate dynamics of the wildfire frequency data, including zero-inflation, time-varying patterns, and the over-dispersed feature compared to the Poisson distribution. Furthermore, the DCMM model is able to incorporate the tie-varying spatial dependence of the frequency data among multiple locations. In the empirical section, we

illustrate the capability of the proposed model to produce accurate mean forecast, as well as reasonable credible intervals for states with rather different patterns (frequent wildfire versus very few wildfires over the sample).

Based on the DCMM model, together with the Bayesian Gaussian dynamic linear model for the claim severity, we price wildfire catastrophe (CAT) bonds with different characteristics. In the hedging analysis, we show that the hedging with wildfire CAT bonds could substantially lower the variability and tail risk of insurers exposed to wildfire risk in different states. A stop-loss reinsurance application and CAT bond hedging application with the county-level data of California are also presented with the proposed model. Although we have focused on the wildfire application in this paper, the proposed Bayesian framework can also be applied to predicting insurance losses of other perils, and study the pricing and hedging effectiveness of CAT bonds written on other peril(s). Such analyses are interesting future research directions. The application of the proposed modeling framework on catastrophe risk capital allocation is another very interesting topic to explore in future research.

References

Berry, Lindsay R., and Mike West. 2020. "Bayesian forecasting of many count-valued time series." *Journal of Business & Economic Statistics* (Taylor & Francis) 38: 872–887.

Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino. 2019. "Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors." *Journal of Econometrics* (Elsevier) 212: 137–154.

Frees, Edward W. 2009. *Regression Modeling with Actuarial and Financial Applications.* Cambridge: Cambridge University Press. doi:10.1017/CBO9780511814372.

Gan, Jianbang, Adam Jarrett, and Cassandra Johnson Gaither. 2014. "Wildfire risk adaptation: propensity of forestland owners to purchase wildfire insurance in the southern United States." *Canadian Journal of Forest Research* (NRC Research Press) 44: 1376–1382.

Hainaut, Donatien, and Jean-Philippe Boucher. 2014. "Frequency and severity modelling using multifractal processes: an application to tornado occurrence in the USA and CAT bonds." *Environmental Modeling & Assessment* (Springer) 19: 207–220.

Kaas, Rob, Marc Goovaerts, Jan Dhaene, and Michel Denuit. 2008. *Modern actuarial risk theory: using R.* Vol. 128. Springer Science & Business Media.

Klugman, Stuart A., Harry H. Panjer, and Gordon E. Willmot. 2012. *Loss Models: From Data to Decisions (4th ed.).* Hoboken: Wiley.

Major, John A., and Rodney E. Kreps. 2002. "Catastrophe risk pricing in the traditional market." *Alternative Risk Strategies* (Risk Waters Group London) 201–222.

McElreath, Richard. 2018. *Statistical rethinking: A Bayesian course with examples in R and Stan.* Chapman and Hall/CRC.

Prado, Raquel, and Mike West. 2010. *Time Series: Modeling, Computation, and Inference.* CRC Press.

Thomas, D., D. Butry, S. Gilbert, D. Webb, and J. Fund. 2017. "The costs and losses of wildfires: a literature survey. US Department of Commerce, National Institute of Standards and Technology."

The costs and losses of wildfires: a literature survey. US Department of Commerce, National Institute of Standards and Technology.

Trottier, Denis-Alexandre, Anne-Sophie Charest, and others. 2018. "CAT bond spreads via HARA utility and nonparametric tests." *The Journal of Fixed Income* (Institutional Investor Journals Umbrella) 28: 75–99.

Wang, Daoping, Dabo Guan, Shupeng Zhu, Michael Mac Kinnon, Guannan Geng, Qiang Zhang, Heran Zheng, et al. 2021. "Economic footprint of California wildfires in 2018." *Nature Sustainability* (Nature Publishing Group) 4: 252–260.

Webb, C., and Eric J. Xu. 2018. "The California wildfire conundrum." Tech. rep., Milliman, San Francisco.

West, Mike, and Jeff Harrison. 2006. *Bayesian Forecasting and Dynamic Models.* Springer, New York.

Zhu, Wenge. 2017. "Wanting robustness in insurance: A model of catastrophe risk pricing and its empirical test." *Insurance: Mathematics and Economics* (Elsevier) 77: 14–23.

About the Authors

Hong Li is an associate professor in the Department of Economics and Finance at the Lang School of Business at the University of Guelph. Hong's research focuses on insurance data analysis and risk management, with applications to longevity, financial and climate risks. Hong has published over 20 papers in leading insurance, actuarial, finance, and economics journals including the *Journal of Risk and Insurance*, *Insurance: Mathematics and Economics*, the *Journal of Banking and Finance*, and the *Journal of Economic Behavior and Organization*. He is a Fellow of the Society of Actuaries and an Associate of the Canadian Institute of Actuaries.

Jianxi Su is an associate professor of actuarial science at Purdue University. His research expertise ranges from mathematics/statistics modeling to diverse applications in insurance. He has published in such reputable actuarial journals as *ASTIN Bulletin*, *Insurance: Mathematics and Economics*, *North American Actuarial Journal*, and *Variance*. Over the past several years, he has successfully completed numerous research projects sponsored by the SOA and CAS. He has ample industrial experience with the ORSA team of Sun Life Financial, Canada.

Appendix

A. Single-Scale Forecasting Results

Table 21 and Table 22 report the 1- to 12-month out-of-sample forecast mean absolute deviance of wildfire frequency obtained from the single-scale DCMM models for states in the Western and the Eastern region, respectively.

Table 21. The Mean Absolute Deviation (MAD) of the Western states over the out-of-sample forecasting horizons k=1,2,...,12 months using single-scale models.

k	AK	AZ	CA	CO	HI	ID	MT	NV
1	0.67	0.45	2.43	1.40	0.22	1.09	3.94	0.52
2	0.67	0.45	2.38	1.39	0.22	1.08	4.13	0.52
3	0.66	0.45	2.40	1.37	0.22	1.08	3.79	0.52
4	0.66	0.45	2.38	1.38	0.22	1.07	3.76	0.51
5	0.65	0.45	2.41	1.39	0.21	1.07	3.78	0.51
6	0.65	0.45	2.43	1.39	0.21	1.06	3.81	0.50
7	0.65	0.46	2.44	1.40	0.21	1.07	3.72	0.50
8	0.64	0.45	2.39	1.38	0.21	1.09	3.82	0.51
9	0.59	0.45	2.37	1.37	0.21	1.09	3.54	0.50
10	0.60	0.45	2.47	1.37	0.21	1.08	3.59	0.50
11	0.59	0.44	2.48	1.35	0.21	1.07	3.39	0.50
12	0.58	0.44	2.45	1.34	0.21	1.06	3.46	0.51
k	NM	OK	OR	TX	UT	WA	WY	
1	1.55	0.77	0.75	11.51	1.84	2.54	0.51	
2	1.52	0.76	0.75	11.11	1.82	2.50	0.51	
3	1.53	0.76	0.73	10.51	1.82	2.52	0.51	
4	1.51	0.76	0.73	10.11	1.80	2.39	0.50	
5	1.51	0.75	0.74	10.11	1.78	2.41	0.50	
6	1.48	0.75	0.74	11.22	1.79	2.31	0.50	
7	1.47	0.75	0.74	9.71	1.78	2.27	0.49	
8	1.48	0.75	0.74	9.41	1.80	2.31	0.49	
9	1.50	0.75	0.74	9.29	1.74	2.28	0.50	
10	1.48	0.75	0.73	8.96	1.76	2.25	0.49	
11	1.47	0.76	0.71	8.78	1.72	2.27	0.49	
12	1.44	0.75	0.70	8.36	1.70	2.31	0.49	

Table 22. The Mean Absolute Deviation (MAD) of the Eastern states over the out-of-sample forecasting horizons k=1,2,...,12 months using single-scale models.

k	AR	FL	GA	KS	LA	MD	MI	MN	MO
1	0.39	1.70	1.27	0.19	0.35	0.15	0.16	0.17	0.47
2	0.39	1.68	1.28	0.19	0.35	0.15	0.15	0.17	0.46
3	0.39	1.67	1.28	0.19	0.35	0.15	0.15	0.17	0.46
4	0.39	1.64	1.29	0.19	0.35	0.15	0.15	0.17	0.45
5	0.39	1.62	1.29	0.19	0.34	0.15	0.15	0.17	0.45
6	0.39	1.63	1.27	0.19	0.34	0.15	0.15	0.17	0.44
7	0.38	1.62	1.28	0.19	0.34	0.14	0.15	0.17	0.45
8	0.39	1.60	1.29	0.19	0.34	0.14	0.15	0.16	0.45
9	0.39	1.61	1.26	0.19	0.34	0.14	0.15	0.16	0.44
10	0.38	1.56	1.25	0.19	0.34	0.14	0.15	0.16	0.44
11	0.38	1.58	1.26	0.19	0.34	0.14	0.15	0.16	0.43
12	0.38	1.53	1.25	0.19	0.33	0.14	0.15	0.16	0.44
k	NE	NJ	NY	NC	SC	VA	WV	WI	
1	0.37	0.18	0.09	0.11	0.12	0.35	1.24	2.48	
2	0.37	0.18	0.09	0.11	0.12	0.35	1.23	2.46	
3	0.37	0.18	0.09	0.11	0.12	0.35	1.23	2.43	
4	0.36	0.18	0.09	0.12	0.12	0.34	1.20	2.31	
5	0.36	0.18	0.09	0.11	0.12	0.35	1.23	2.39	
6	0.36	0.18	0.09	0.12	0.12	0.34	1.19	2.31	
7	0.36	0.18	0.09	0.12	0.12	0.33	1.16	2.29	
8	0.36	0.18	0.08	0.11	0.12	0.34	1.21	2.23	
9	0.35	0.17	0.08	0.11	0.12	0.34	1.16	2.23	
10	0.35	0.17	0.08	0.11	0.11	0.34	1.19	2.22	
11	0.36	0.17	0.09	0.11	0.11	0.33	1.14	2.18	
12	0.36	0.17	0.08	0.11	0.11	0.33	1.14	2.14	