Check for updates

The Canadian Journal of Statistics | La revue canadienne de statistique
Vol. 54, No. 1, 2026, e70024

An observation-driven state-space model for claims size modelling

Jae Youn AHN
Department of Statistics, Ewha Womans University, Seoul, Republic of Korea

Himchan JEONG @
Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, British Columbia, Canada

Mario V. WÜTHRICH
Department of Mathematics, ETH Zurich, Zurich, Switzerland

Abstract State-space models are popular in econometrics. Recently, these models have gained some popularity in the actuarial literature. The best known state-space models are of the Kalman-filter type. These are called parameter-driven because the observations do not impact the state-space dynamics. A second less well-known class of state-space models comprises the so-called observation-driven state-space models where the state-space dynamics is also impacted by the actual observations. A typical example is the Poisson-gamma observation-driven state-space model for count data, which is fully analytically tractable. The goal of this article is to develop a gamma-gamma observation-driven state-space model for claim size modelling. We provide fully tractable versions of gamma-gamma observation-driven state-space models; these versions extend the work of the Smith–Miller model by allowing for a fully flexible variance behaviour. Additionally, we demonstrate that the proposed model aligns with evolutionary credibility, a methodology in insurance that dynamically adjusts premium rates over time using evolving data.

Résumé Les modèles espace-état occupent une place prépondérante en économétrie et suscitent un intérêt croissant dans la littérature actuarielle. Parmi ces modèles, ceux de type filtre de Kalman, dits pilotés par les paramètres, sont les plus répandus : leur dynamique d'état n'est pas influencée par les observations. Une seconde catégorie, moins répandue, regroupe les modèles espace-état pilotés par les observations, pour lesquels la dynamique de l'espace d'état dépend également des réalisations observées. Le modèle espace-état Poisson-gamma piloté par les observations pour données de comptage constitue un exemple typique, ce modèle étant entièrement traitable analytiquement. Cet article propose le développement d'un modèle espace-état gamma-gamma piloté par les observations, spécifiquement conçu pour la modélisation de la taille des sinistres. Les auteurs y présentent des versions entièrement traitables de ce modèle, étendant ainsi les travaux de Smith et Miller en autorisant une structure de variance pleinement flexible. Par ailleurs, il est démontré que le modèle proposé s'intègre naturellement dans le cadre de la crédibilité évolutive, une méthodologie actuarielle permettant l'ajustement dynamique des tarifs de primes à partir de données évolutives.

Keywords Claim size; evolutionary credibility; Kalman filter; observation-driven state-space model.
MSC2020 Primary 62P05; Secondary 62M10.
@ Corresponding author himchanj@sfu.ca

© 2025 The Author(s). The Canadian Journal of Statistics | La revue canadienne de statistique published by Wiley Periodicals LLC on behalf of Statistical Society of Canada | Société statistique du Canada.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/cjs.70024

1 Introduction

Static random-effects models (Laird and Ware, 1982; Lee and Nelder, 1996) have a long tradition in insurance ratemaking to account for heterogeneity among insurance policyholders (Bichsel, 1964). These models assume that individual claim behaviour is described by an unobserved latent variable that remains constant over time for a given policyholder. By inferring this latent variable from past observations, static random-effects models provide a systematic way of predicting future claims based on past claims experience. This approach has proven effective in contexts where temporal dynamics are less critical, as it simplifies the modelling process and requires fewer assumptions about the evolution of risk factors.

However, in many real-world applications, it is crucial to capture temporal dynamics and evolving risks, necessitating an extension from static random-effects models to dynamic frameworks. State-space models (Kalman, 1960; Anderson and Moore, 2005) naturally comply with this need by introducing latent processes that evolve over time, allowing the modelling of longitudinal data with temporal dependencies. These models are particularly suitable for insurance applications, where the evolution of risk factors plays a critical role (Pinquet et al., 2001; Bolancé et al., 2003). Parameter-driven state-space models, which rely on stochastic latent processes, provide a robust framework for such settings. However, these models often lack analytical tractability, as closed-form expressions for filtering and likelihood evaluations are generally unavailable except in limited cases, such as under Gaussian assumptions. As a result, numerical methods are employed, which increase computational complexity, require significant processing time, and may restrict explainability because sensitivity analyses may not be available easily. This makes practical implementation challenging in large-scale or real-time applications (Doucet et al., 2000; Arulampalam et al., 2002).

Alternatively, credibility methods (Whitney, 1918; Bühlmann, 1967; Bühlmann and Straub, 1970) can be used as a simpler (approximative) solution to predict claims. Parameter-driven state-space versions were studied in Pinquet (2020a, 2020b) and Ahn et al. (2021). However, credibility methods typically focus on providing only the predictive mean, and they do not yield the full predictive distribution. Of course, this is a significant limitation of credibility methods, e.g., in risk management where the full predictive distributions need to be studied.

To address these challenges, observation-driven state-space models (Smith and Miller, 1986; Harvey and Fernandes, 1989) have proven to be a practical and effective alternative. These models maintain the dynamic structure of state-space frameworks while allowing for analytical solutions through the use of conjugate prior relationships. Smith and Miller (1986) introduced an observation-driven state-space model for exponentially distributed responses, which was later applied to insurance ratemaking (Bolancé et al., 2007). This model was subsequently extended to gamma-distributed responses using the Bayesian gamma-gamma conjugacy (Ahn et al., 2023a), increasing its relevance and applicability in insurance modelling. Similarly, Harvey and Fernandes (1989) introduced an observation-driven state-space model tailored for count data.

Despite their advantages, existing observation-driven state-space models (Smith and Miller, 1986; Harvey and Fernandes, 1989) are limited in their ability to capture diverse variance behaviours. Specifically, these models typically assume an increasing variance process that asymptotically goes to infinity. This restricts their flexibility in applications where the variance dynamics may vary, may be uniformly bounded, or may even be decreasing. The limitations of the model of Harvey and Fernandes (1989) were formally analyzed in Ahn et al. (2023b), where flexible variance extensions were proposed to address these constraints. The model of Smith and Miller (1986) has not yet been evaluated in this regard, leaving its ability to handle diverse variance behaviours unexplored. In this article, we address this gap by examining the limitations of the model of Smith and Miller (1986) and providing a generalized framework that overcomes these issues.

To this end, we extend the model introduced in Smith and Miller (1986) to develop a flexible observation-driven state-space framework specifically designed for continuous positive data, based on gamma distributions. These generalized models are able to accommodate the full range of variance behaviours, including stationary, increasing, and decreasing variance patterns, thereby addressing the constraints of the original framework. By retaining the analytical tractability of the original observation-driven approach, the proposed models ensure closed-form solutions for filtering, estimation, and predictive distributions while significantly enhancing the flexibility in modelling the variance dynamics. These features make these models particularly well suited for applications involving continuous positive data, such as insurance claim sizes. We acknowledge that in order to model the full insurance claim process of a policyholder (or any risk exposure unit), one additionally needs a claims frequency risk model to arrive at the compound claims process. This is not included in this article but we refer to the actuarial literature for claims frequency modelling; for an observation-driven version,

1708945x, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjs.70024 by Ewha Womans University Library, Wiley Online Library on [21/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

see Ahn et al. (2025). Furthermore, we show that the proposed model is consistent with evolutionary credibility, dynamically adjusting to evolving claims data for application in insurance pricing schemes.

This article is structured as follows. We first revisit the Smith–Miller model, paying particular attention to the variance behaviour of its state-space process. To achieve full flexibility in the variance behaviour, we introduce the Generalized Smith–Miller model in Section 3 and analyze its variance dynamics in Section 4. Section 5 covers model fitting, forecasting, and the evolutionary credibility formula within the Generalized Smith–Miller framework. Section 6 and Appendix A present a real data example and a simulation study, respectively. Finally, concluding remarks are provided in Section 7.

2 Revisiting the Smith–Miller model

We start by revisiting the Smith–Miller model (Smith and Miller, 1986) which originally utilized the exponential distribution to characterize the observation distribution. This framework was later extended (Ahn et al., 2023a) to include the more general gamma distribution, allowing additionally for over-dispersion. We emphasize the need to generalize this latter model by demonstrating that it can accommodate only a very limited variance behaviour of the state-space dynamics.

We begin by introducing the notation. For finite sequences, let $Y_{1:t} = (Y_1, \dots, Y_t)$, and interpret $Y_{1:0}$ as the empty sequence, which generates the trivial σ-field $\sigma(Y_{1:0}) = \{\emptyset, \Omega\}$ on the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We denote by $\gamma(\alpha, \beta)$ the gamma distribution with shape parameter $\alpha > 0$ and rate parameter $\beta > 0$. It is positively supported and has probability density function on $\mathbb{R}_+ := \{x | x > 0, x \in \mathbb{R}\}$

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\gamma(\alpha)} \, x^{\alpha-1} \exp(-\beta x) \qquad \text{for } x > 0.$$

The mean and variance are given by α/β and α/β^2, respectively.

Typically, in insurance applications, claim severities Y are observed in a longitudinal format, resulting in doubly indexed claims severities $Y_{i,t}$—for policyholder i in time period t. This is the typical structure of longitudinal insurance data. Usually, one assumes that the individual claims $Y_{i,t}$ of different policyholders are independent and, to not overload the notation, we omit the policyholder index i when introducing the Generalized Smith–Miller model, hence, interpreting $(Y_t)_{t\geq 1}$ as time-series data. In this regard, the policyholder index i becomes relevant only when the proposed framework is applied to experience rating of policyholders where we study the log-likelihood function simultaneously over all available policyholders $1 \leq i \leq M$, see Section 6 and Appendix A.

2.1 Review of the Smith–Miller model

The Smith–Miller model involves two stochastic processes: a latent state-space process $(\Theta_t)_{t\geq 1}$, which acts as a hidden driver of the dynamics, and an observable process of response variables $(Y_t)_{t\geq 1}$, which depends on the latent state-space process. A key feature of the model is the incorporation of a feedback loop from the responses to the state-spaces, impacting the state-space updates; see the specification of the distribution (4), below. This feedback mechanism distinguishes observation-driven state-space models from their parameter-driven counterparts of Kalman-filter type (Kalman, 1960). Specifically, following the classification in Cox (1981), parameter-driven state-space models are defined by specifying the state-space updates

$$\text{from} \quad \Theta_t \big| \Theta_{1:t-1} \quad \text{to} \quad \Theta_{t+1} \big| \Theta_{1:t}.$$

In contrast, in observation-driven models, the state-space update can be defined by specifying the state-space transitions

$$\text{from} \quad \Theta_t | Y_{1:t} \quad \text{to} \quad \Theta_{t+1} | Y_{1:t}.$$

See Remark 1 for the motivation of this transition. This incorporates the observed responses into the state-space update.

Smith–Miller model (Smith and Miller, 1986; Ahn et al., 2023a). Consider a fixed dispersion parameter $\psi > 0$, initialization $a_{1|0} > 1$, and two exogenous sequences $(v_t)_{t\geq 1}$ and $(\mu_t)_{t\geq 1}$, where $v_t \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$ and $\mu_t \in \mathbb{R}_+$. The processes of the responses $(Y_t)_{t\geq 1}$ and of the state-spaces $(\Theta_t)_{t\geq 1}$ are defined as follows:

(0) *Initialization at* $t = 1$: Initialize the state-space process $(\Theta_t)_{t\geq 1}$ by

$$\Theta_1 | Y_{1:0} \sim \gamma\big(1 + a_{1|0}, b_{1|0}\big) \text{ with } a_{1|0} = b_{1|0} > 0, \tag{1}$$

giving unit mean of the inverse initial state, $\mathrm{E}[\Theta_1^{-1}] = 1$.

1708945x, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjs.70024 by Ewha Womans University Library, Wiley Online Library on [21/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(1) *Observation equation for* $t \geq 1$: For given $Y_{1:t-1}$ and $\Theta_{1:t}$, the response Y_t satisfies

$$Y_t|Y_{1:t-1}, \Theta_{1:t} \sim \gamma\Big(\frac{\upsilon_t}{\psi}, \frac{\Theta_t}{\mu_t \psi}\Big), \quad \text{if } \upsilon_t > 0, \tag{2}$$

and $Y_t|Y_{1:t-1}, \Theta_{1:t} = 0$, almost surely, if $\upsilon_t = 0$.

(2) *State-space update from* t *to* $t+1$: Assume the filtering distribution at time $t \geq 1$ for positive constants a_t and b_t is given by

$$\Theta_t|Y_{1:t} \sim \gamma(1 + a_t, b_t). \tag{3}$$

Then, the predictive distribution for Θ_{t+1} is assumed to satisfy

$$\Theta_{t+1}|Y_{1:t} \sim \gamma\big(1 + a_{t+1|t}, b_{t+1|t}\big), \tag{4}$$

where $a_{t+1|t}$ and $b_{t+1|t}$ follow the updating rule

$$a_{t+1|t} = q_t a_t > 1 \qquad \text{and} \qquad b_{t+1|t} = q_t b_t > 0,$$

with q_t being defined by

$$q_t = \frac{\gamma(a_t - 1) + 1}{a_t}, \tag{5}$$

for a constant parameter $\gamma \in (0, 1]$.

Remark 1. We show in (14) below that the distribution (3) is a straightforward implication of Bayes's rule. Together with (5) and the initial condition $a_{1|0} > 1$, this implies that $a_t > 1$ as well as $a_{t|t-1} > 1$, for all $t \geq 1$, so that the conditional variances of the reciprocals of the states are well defined for all $t \geq 1$. We also discuss the conditional distribution of the observations (2) and its application to insurance. For $\upsilon_t \geq 1$, it provides us with the first two conditional moments

$$\mathrm{E}[\,Y_t|Y_{1:t-1}, \Theta_{1:t}] = \frac{\upsilon_t \mu_t}{\Theta_t} \qquad \text{and} \qquad \mathrm{var}[\,Y_t|Y_{1:t-1}, \Theta_{1:t}] = \upsilon_t \psi\, \frac{\mu_t^2}{\Theta_t^2}. \tag{6}$$

Here, μ_t is the component of the scale parameter that may depend on policyholder covariates (information) to capture the observable heterogeneity, whereas $1/\Theta_t$ is a random effect, i.e., the counterpart of μ_t that reflects the unobservable heterogeneity in the distributions. For a non-unit exposure (which means the extent to which a business or individual is susceptible to various perils) $\upsilon_t \geq 2$, Y_t can be interpreted as the sum of υ_t conditionally i.i.d. gamma distributed individual claims, that is,

$$Z_{t,1}, \ldots, Z_{t,\upsilon_t}\big|\Theta_t \overset{\text{i.i.d.}}{\sim} \gamma(1/\psi, \Theta_t/(\mu_t \psi)),$$

since we have the distributional identity

$$Y_t|Y_{1:t-1}, \Theta_{1:t} \overset{\text{d}}{=} \sum_{j=1}^{\upsilon_t} Z_{t,j}\big|\Theta_t \sim \gamma\Big(\frac{\upsilon_t}{\psi}, \frac{\Theta_t}{\mu_t \psi}\Big). \tag{7}$$

Note that the models in Smith and Miller (1986) and Ahn et al. (2023a) do not include exposures υ_t in the observation equation (2), but for insurance modelling this is very convenient to collect multiple i.i.d. claims within the same period t. It is sufficient to know the aggregate (or average) claim amount Y_t for each time period t, because it forms a sufficient statistic for parameter estimation in this problem.

Lastly, in describing the Smith–Miller model, we omit the parametric specification of the state-space update from (3) to (4). The above model assumptions are thus not complete. Similar to Smith and Miller (1986) and Harvey and Fernandes (1989), such a specification is readily available, as shown in Lemma 6 in Appendix C. However, we note that parts ii and iv of Lemma 6 reveal that these specifications are not unique. Furthermore, interestingly, the specific choice of the joint distribution between Θ_t and Θ_{t+1}, given $Y_{1:t}$, does not influence the predictive model, provided the conditional distribution of Θ_{t+1}, given $Y_{1:t}$, is preserved. This follows because

1708945x, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjs.70024 by Ewha Womans University Library, Wiley Online Library on [21/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the predictive density of Y_{t+1}, given $Y_{1:t}$, relies only on the marginal distribution of Θ_{t+1}, as the following computation shows:

$$\begin{aligned} f(Y_{t+1} \mid Y_{1:t}) &= \int f(Y_{t+1}, \Theta_{1:t+1} \mid Y_{1:t})\, d\Theta_{1:t+1} \\ &= \int f(Y_{t+1} \mid \Theta_{t+1})\, f(\Theta_{1:t+1} \mid Y_{1:t})\, d\Theta_{1:t+1} \\ &= \int f(Y_{t+1} \mid \Theta_{t+1})\, f(\Theta_{t+1} \mid Y_{1:t})\, d\Theta_{t+1}, \end{aligned} \tag{8}$$

where the second equality uses the conditional independence from specification (2). Thus, the copula between Θ_t and Θ_{t+1}, given $Y_{1:t}$, can be chosen arbitrarily to predict Y_{t+1}, as long as the marginal distribution of Θ_{t+1}, given $Y_{1:t}$, remains unchanged; in fact, (8) shows that the specific copula choice is integrated out by this marginalization.

The state-space update under the Smith–Miller model displays a mean-stationary dynamics of the inverse state-space process $(\Theta_t^{-1})_{t\geq 1}$ when passing from filtering to the predictive distribution, i.e.,

$$\mathrm{E}[\,\Theta_{t+1}^{-1}\big|Y_{1:t}] = \mathrm{E}[\,\Theta_t^{-1}\big|Y_{1:t}]. \tag{9}$$

This in turn implies unconditional mean stationarity of the inverse state-space process. On the other hand, the conditional variances of the inverse state-spaces are *proportionally increasing*

$$\mathrm{var}[\,\Theta_{t+1}^{-1}\big|Y_{1:t}] = \frac{1}{\gamma}\,\mathrm{var}[\,\Theta_t^{-1}\big|Y_{1:t}]. \tag{10}$$

Note that in analyzing the "stationarity of the state-spaces," we are mainly interested in the mean and the variance stationarity of the inverse state-space process $(\Theta_t^{-1})_{t\geq 1}$ rather than the process $(\Theta_t)_{t\geq 1}$ itself. This distinction is motivated by the conditional mean and variance in (6), which highlights the primary interest in the mean and variance behaviour of $(\Theta_t^{-1})_{t\geq 1}$.

The popularity of the Smith–Miller model is explained by the fact that it allows for explicit formulations of the predictive means $\mathrm{E}[Y_{t+1}|Y_{1:t}]$ and the probability density functions of the response variables $Y_{1:t}$ in a recursive form (Ahn et al., 2023a). The purpose of our generalization is to allow for more modelling flexibility in the variance behaviour (10), as further discussed in the next section.

2.2 Increasing variance behaviour of the inverse state-space dynamics

The Smith–Miller model enjoys considerable popularity in insurance and other applied fields due to its intuitive interpretation of state-space dynamics, as demonstrated in equations (9) and (10), along with its analytically tractable likelihood function. Nevertheless, a notable limitation arises from the restrictive assumption about the variance dynamics of the inverse state-space process. Specifically, the Smith–Miller model inherently constrains the variance dynamics to an increasing pattern, precluding other plausible behaviours such as stationary or decreasing variance. The following result formally clarifies this limitation.

Proposition 1. *Consider the setting of the Smith–Miller model. Then we have*

$$\mathrm{var}[\Theta_t^{-1}] \leq \mathrm{var}[\mathrm{E}[\Theta_t^{-1} \mid Y_{1:t}]] + \frac{1}{\gamma}\,\mathrm{E}\Big[\mathrm{var}[\Theta_t^{-1} \mid Y_{1:t}]\Big] = \mathrm{var}[\Theta_{t+1}^{-1}] < \infty, \quad t \geq 1$$

with a strict inequality for $\gamma \in (0,1)$.

Proof of Proposition 1. For $\gamma \in (0,1]$, the tower property for conditional expectations gives us

$$\begin{aligned} \mathrm{var}[\Theta_t^{-1}] &= \mathrm{var}[\mathrm{E}[\Theta_t^{-1} \mid Y_{1:t}]] + \mathrm{E}\Big[\mathrm{var}[\Theta_t^{-1} \mid Y_{1:t}]\Big] \\ &\leq \mathrm{var}[\mathrm{E}[\Theta_t^{-1} \mid Y_{1:t}]] + \frac{1}{\gamma}\,\mathrm{E}\Big[\mathrm{var}[\Theta_t^{-1} \mid Y_{1:t}]\Big] \\ &= \mathrm{var}[\mathrm{E}[\Theta_{t+1}^{-1} \mid Y_{1:t}]] + \mathrm{E}\Big[\mathrm{var}[\Theta_{t+1}^{-1} \mid Y_{1:t}]\Big] = \mathrm{var}[\Theta_{t+1}^{-1}], \end{aligned}$$

17089345x, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjs.70024 by Ewha Womans University Library, Wiley Online Library on [21/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where the last equality follows from (9) and (10). Furthermore, since $\mathrm{E}[\mathrm{var}[\Theta_t^{-1}|Y_{1:t}]] > 0$, the equality becomes strict for $\gamma \in (0,1)$. This completes the proof. □

This proposition demonstrates that, under the Smith–Miller model, the variance of the inverse state-space process exhibits strictly increasing behaviour over time for the case $\gamma < 1$. For insurance modelling, one may require a different behaviour, e.g., one may argue that with increasing time (and information) uncertainty should be decreasing. Such an extension (generalization) of the Smith–Miller model is introduced in the next section.

3 Generalized Smith–Miller model

Proposition 1 demonstrates that the state-space updates (4) cannot lead to a decreasing or stationary variance behaviour of the inverse state-space dynamics. We introduce a generalized version of the Smith–Miller model. Modifying the state-space update (4) appropriately will allow for flexible variance behaviours, whose detailed analysis is provided in Section 4.

Generalized Smith–Miller model. Consider a fixed dispersion parameter $\psi > 0$, initialization $a_{1|0} > 0$, and three exogenous sequences $(\xi_t)_{t\geq 1}$, $(v_t)_{t\geq 1}$, and $(\mu_t)_{t\geq 1}$, where

$$\xi_t \in \Xi, \quad v_t \in \mathbb{N}_0, \quad \mu_t \in \mathbb{R}_+,$$

and $\Xi \subseteq \mathbb{R}^K$ is a convex parameter space for a positive integer K. The processes of the responses $(Y_t)_{t\geq 1}$ and the state-spaces $(\Theta_t)_{t\geq 1}$ are assumed to have the same initialization (1), $t = 1$, and the same observation equation (2), $t \geq 1$, as the Smith–Miller model. The state-space update from time t to time $t+1$ is defined as follows:

(2) *State-space update from* t to $t+1$: Assume that the filtering distribution at time $t \geq 1$ is given by

$$\Theta_t|Y_{1:t} \sim \gamma(1 + a_t, b_t), \tag{11}$$

where a_t and b_t are positive functions of $Y_{1:t}$. Then the predictive distribution for Θ_{t+1} is assumed to satisfy

$$\Theta_{t+1}|Y_{1:t} \sim \gamma\big(1 + a_{t+1|t}, b_{t+1|t}\big), \tag{12}$$

where $a_{t+1|t}$ and $b_{t+1|t}$ follow the recursion

$$a_{t+1|t} = A(a_t, b_t; \xi_t) > 0 \qquad \text{and} \qquad b_{t+1|t} = B(a_t, b_t; \xi_t) > 0, \tag{13}$$

for given measurable functions $A : \mathbb{R}_+^2 \times \Xi \to \mathbb{R}_+$ and $B : \mathbb{R}_+^2 \times \Xi \to \mathbb{R}_+$ specified below.

With suitable choices for the update functions A and B in equation (13), the model is fully tractable with a mean stationary state-space process. Section 4 examines the long-term variance behaviour of the inverse state-space process $(\Theta_t^{-1})_{t\geq 1}$. Before turning to its variance properties, we first discuss in the present section the mean stationarity condition of the process. This discussion is given in detail below. Before proceeding to this, we briefly address the well-definedness of the Generalized Smith–Miller model. Although we introduced formula (11) as an assumption, it is, in fact, a direct consequence of the conditional response distribution (2) and the latent distribution (12). This result is obtained by applying a Bayesian inference step, often referred to as filtering in the time-series literature, within a gamma-gamma conjugate prior framework. To clarify this, we briefly explain the process, generically denoting a density by f. By applying Bayes's formula

$$f(\Theta_t|Y_{1:t}) \propto f(Y_t|Y_{1:t-1}, \Theta_t) f(\Theta_t|Y_{1:t-1}),$$

and substituting the explicit forms of the gamma densities, the filtering distribution in equation (11) becomes a gamma distribution with the following parameter updates:

$$a_t := a_{t|t-1} + \frac{v_t}{\psi} > 0 \qquad \text{and} \qquad b_t := b_{t|t-1} + \frac{Y_t}{\mu_t \psi} > 0, \tag{14}$$

where $(a_{t|t-1}, b_{t|t-1})$ and $(v_t/\psi, Y_t/\mu_t\psi)$ are from (12) and (2), respectively. In this sense, formula (11) is not a model assumption but rather a (mathematical) consequence of the state-space update in (12), the conditional response distribution (2), and the initial condition (1). As a result, the Generalized Smith–Miller model is well defined.

1708945x, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjs.70024 by Ewha Womans University Library, Wiley Online Library on [21/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

We now discuss the state-space update (12) and we specify explicit functional forms of A and B in equation (13) that result in linearly tractable models with a mean stationary inverse state-space dynamics. Among various possible choices of the update functions A and B, we consider the following positive affine functions for $\xi_t = (\xi_{1,t}, \ldots, \xi_{6,t})^\top \in \Xi$

$$a_{t+1|t} = A(a_t, b_t; \xi_t) = \xi_{1,t} + \xi_{2,t} a_t + \xi_{3,t} b_t > 0,$$

and

$$b_{t+1|t} = B(a_t, b_t; \xi_t) = \xi_{4,t} + \xi_{5,t} a_t + \xi_{6,t} b_t > 0.$$

If, for some $\Delta_t \in (0, 1]$, we furthermore require the following thinning property

$$\mathrm{E}\left[\Theta_{t+1}^{-1} \mid Y_{1:t}\right] = \Delta_t \mathrm{E}\left[\Theta_t^{-1} \mid Y_{1:t}\right] + (1 - \Delta_t), \qquad \text{for all } a_t, b_t > 0, \tag{15}$$

then this implies the restriction for $t \geq 1$

$$\xi_{1,t} = \xi_{3,t} = \xi_{4,t} = 0 \qquad \text{and} \qquad \xi_{6,t} = \Delta_t \xi_{2,t}.$$

The thinning in (15) is motivated by the intuition that we would like to carry forward some past information $Y_{1:t}$, reflected by the first term on the r.h.s. in equation (15), but we also want to add some new noise to the state-space process reflected by the second term $(1 - \Delta_t) \cdot 1$, where this can be interpreted as the contribution of the prior mean part $\mathrm{E}[\Theta_t^{-1}] = 1$. The validity of this prior mean is implied by mean stationarity, which still needs to be established; see Proposition 2, below. After an appropriate reparameterization, we present the following linear evolutionary model as a specific instance of the Generalized Smith–Miller model.

Model 1. Consider a fixed dispersion parameter $\psi > 0$, initialization $a_{1|0} > 0$, and four exogenous sequences $(p_t)_{t\geq 1}$, $(q_t)_{t\geq 1}$, $(\upsilon_t)_{t\geq 1}$, and $(\mu_t)_{t\geq 1}$, where

$$p_t \in [0, 1], \quad q_t \in \mathbb{R}_+, \quad \upsilon_t \in \mathbb{N}_0 \quad \text{and} \quad \mu_t \in \mathbb{R}_+.$$

Consider the processes of the responses $(Y_t)_{t\geq 1}$ and the state-spaces $(\Theta_t)_{t\geq 1}$ as in the Generalized Smith–Miller model, where the general recursions (13) are replaced by, for $t \geq 1$,

$$a_{t+1|t} = (p_t + q_t) a_t \qquad \text{and} \qquad b_{t+1|t} = p_t a_t + q_t b_t. \tag{16}$$

Model 1 is well defined if $a_{t+1|t} > 0$ and $b_{t+1|t} > 0$ for all $t \geq 0$, which is an immediate consequence of recursions (14) and (16) as well as the initialization $a_{1|0} = b_{1|0} > 0$. Simple algebraic computations show that Δ_t in (15) can be represented as

$$\Delta_t = q_t / (p_t + q_t) \in (0, 1]. \tag{17}$$

Hence, the pair of exogenous sequences $(p_t)_{t\geq 1}$ and $(q_t)_{t\geq 1}$ in Model 1 can be equivalently represented by $(q_t)_{t\geq 1}$ and the credibility weights $(\Delta_t)_{t\geq 1}$ via the relation in (17). In view of (14), along with the updates in (16), the observations $(Y_t)_{t\geq 1}$ only enter the sequence $(b_{t+1|t})_{t\geq 1}$, and, consequently, $(a_{t+1|t})_{t\geq 1}$ is deterministic.

Remark 2. Parametric representations of the state-space updates, such as the one from (11) to (12), have been highlighted as the key steps in the related literature (Smith and Miller, 1986; Harvey and Fernandes, 1989; Ahn et al., 2023a, 2025). These representations are often based on thinning a (gamma) process using the techniques presented in Appendix C. This leads to a full stochastic description of the corresponding observation-driven state-space model. Based on Lemma 6, we could present a similar explicit parametric representation for model 1. However, it is important to realize that such parametric representations are not unique if one is only interested in the conditional distributions of the responses Y_{t+1}, given $Y_{1:t-1}$, see equation (6). This non-uniqueness diminishes the interpretive value of these constructions (if only $(Y_t)_{t\geq 1}$ is observable), as different parameterizations (and copulas) may result in the same underlying observation dynamics.

Proposition 2. *Under the setting of Model* 1, *we have mean stationarity* $\mathrm{E}[\Theta_t^{-1}] = 1$, $t \geq 1$.

1708945x, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjs.70024 by Ewha Womans University Library, Wiley Online Library on [21/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Proof of Proposition 2. We compute

$$\mathrm{E}\Big[\Theta_{t+1}^{-1}\Big|Y_{1:t}\Big] = \frac{b_{t+1|t}}{a_{t+1|t}} = \frac{q_t b_t + p_t a_t}{(p_t+q_t)a_t} = \frac{q_t}{p_t+q_t}\frac{b_t}{a_t} + \frac{p_t}{p_t+q_t}$$
$$= \Delta_t \mathrm{E}\Big[\Theta_t^{-1}\Big|Y_{1:t}\Big] + (1-\Delta_t). \tag{18}$$

Using the tower property of conditional expectations, this implies

$$\mathrm{E}\Big[\Theta_{t+1}^{-1}\Big|Y_{1:t-1}\Big] = \mathrm{E}\Big[\mathrm{E}\Big[\Theta_{t+1}^{-1}\Big|Y_{1:t}\Big]\Big|Y_{1:t-1}\Big] = \Delta_t \mathrm{E}\Big[\Theta_t^{-1}\Big|Y_{1:t-1}\Big] + (1-\Delta_t)$$
$$= \Delta_t\Delta_{t-1}\mathrm{E}\Big[\Theta_{t-1}^{-1}\Big|Y_{1:t-2}\Big] + (1-\Delta_t\Delta_{t-1}) = \cdots = \left(\prod_{k=1}^{t}\Delta_k\right)\mathrm{E}\Big[\Theta_1^{-1}\Big|Y_{1:0}\Big] + \left(1-\prod_{k=1}^{t}\Delta_k\right).$$

Since $\mathrm{E}\Big[\Theta_1^{-1}|Y_{1:0}\Big] = \mathrm{E}\Big[\Theta_1^{-1}\Big] = b_{1|0}/a_{1|0}$ and $a_{1|0} = b_{1|0}$ by the initialization, we have $\mathrm{E}\Big[\Theta_t^{-1}\Big] = 1$ for all $t \geq 1$. □

4 Variance behaviour of the inverse state-space process

In Proposition 2, we proved that the process of the inverse states $(\Theta_t^{-1})_{t\geq 1}$, which is called the "inverse state-space process", is mean stationary under model 1.

The goal of this section is to analyze its long-term variance behaviour under different specifications of the exogenous sequences $(p_t)_{t\geq 1}$ and $(q_t)_{t\geq 1}$ with $p_t \in [0,1]$ and $q_t \in \mathbb{R}_+$ for $t \geq 1$.

We start with a general result, Lemma 1, which is the basic tool to analyze the variance behaviours in the subsequent models and subsections: in Section 4.2, we give parameter sequences that lead to a variance stationary inverse state-space process; in Section 4.3, we provide an example with an increasing variance behaviour that can asymptotically explode (diverge); and in Section 4.4, we provide an example that has a decreasing variance behaviour, in the extreme case converging to zero. That is, asymptotically there is no randomness coming from the state-space process.

4.1 A general result on the variance of the inverse state-space process

We start with a general recursive formula for the variance of the inverse state-space process.

Lemma 1. *Consider the setting of Model 1, and assume that the variances are finite*

$$\mathrm{var}[\Theta_t^{-1}] < \infty, \qquad \text{for all } t \geq 1,$$

which is equivalent to $a_t > 1$. Then the variances of the inverse state-space process satisfy the following recursion:

$$\mathrm{var}[\Theta_{t+1}^{-1}] = \frac{q_t^2}{p_t+q_t}\,\frac{a_t-1}{(p_t+q_t)a_t-1}\mathrm{var}[\Theta_t^{-1}] + \left(1-\frac{q_t^2}{p_t+q_t}\right)\frac{1}{(p_t+q_t)a_t-1}. \tag{19}$$

Proof of Lemma 1. See Appendix B. □

4.2 Model with a stationary variance process

We start with the stationary variance process case under model 1. Recall the weights $\Delta_t = q_t/(p_t+q_t) \in (0,1]$. Moreover, the sequence $(a_t)_{t\geq 1}$ is deterministic, i.e., it does not depend on the responses $(Y_t)_{t\geq 1}$; these only enter the sequence $(b_t)_{t\geq 1}$.

Lemma 2. *Consider the setting of Model 1 with the further condition $a_{1|0} > 1$, and assume that the variances of the inverse state-space process are finite*

$$\mathrm{var}[\Theta_t^{-1}] < \infty, \qquad \text{for all } t \geq 1.$$

For the sequence $(\mathrm{var}(\Theta_t^{-1}))_{t\geq 1}$ to remain constant over time, the exogenous sequences $(p_t)_{t\geq 1}$ and $(q_t)_{t\geq 1}$—or equivalently $(\Delta_t)_{t\geq 1}$ and $(q_t)_{t\geq 1}$—need to satisfy, for all $t \geq 1$,

$$q_t = \frac{\Delta_t a_{1|0}}{a_t - \Delta_t^2 a_t + \Delta_t^2 a_{1|0}}. \tag{20}$$

1708945x, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjs.70024 by Ewha Womans University Library, Wiley Online Library on [21/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Proof of Lemma 2. Lemma 1 implies that for a constant variance process $(\mathrm{var}(\Theta_t^{-1}))_{t\geq 1}$, we need to have for all $t \geq 1$ the identity

$$\frac{1}{a_{1|0}-1} = \frac{q_t^2}{p_t+q_t}\,\frac{a_t-1}{(p_t+q_t)a_t-1}\left(\frac{1}{a_{1|0}-1}\right)+\left(1-\frac{q_t^2}{p_t+q_t}\right)\frac{1}{(p_t+q_t)a_t-1}$$
$$= q_t\Delta_t\,\frac{a_t-1}{q_t a_t/\Delta_t-1}\left(\frac{1}{a_{1|0}-1}\right)+(1-q_t\Delta_t)\frac{1}{q_t a_t/\Delta_t-1},$$

which uses the initialization $\mathrm{var}(\Theta_1^{-1}) = 1/(a_{1|0}-1)$, $a_{1|0}>1$. The above identity is equivalent to

$$q_t a_t/\Delta_t - 1 = q_t\Delta_t\,(a_t-1)+(1-q_t\Delta_t)(a_{1|0}-1) = q_t\Delta_t\,(a_t-a_{1|0})+a_{1|0}-1>0.$$

Solving this for q_t gives the result. □

Lemma 2 is void unless there exists a Model 1 where the corresponding variance process $(\mathrm{var}(\Theta_t^{-1}))_{t\geq 1}$ is constant over time $t \geq 1$. The following explicit example of Model 1 demonstrates that the condition of constant variance in Lemma 2 can indeed be satisfied, thereby ensuring the lemma's applicability and practical relevance. The subsequent result demonstrates that this model exhibits the variance stationarity property.

Model 2. With the additional exogenous sequence

$$(\Delta_t)_{t\geq 1} \qquad \text{with } \Delta\in(0,1],$$

consider the setting in Model 1 with further the condition $a_{1|0}>1$, and where the sequence $(q_t)_{t\geq 1}$ is defined by (20).

For model 2 to be well defined, we need to show $q_t \in \mathbb{R}_+$ for all $t \geq 1$. For this, we show recursively that $q_t > 0$ and $a_t > 0$, assuming that $q_{t-1} > 0$ and $a_{t-1} > 0$. The second sequence is initialized by $a_1 \geq a_{1|0} > 1$, and the first one by

$$q_1 = \frac{\Delta_1 a_{1|0}}{a_1(1-\Delta_1^2)+\Delta_1^2 a_{1|0}} > 0. \tag{21}$$

The recursive step for a_t follows from

$$a_t \geq a_{t|t-1} = (p_{t-1}+q_{t-1})a_{t-1} = q_{t-1}\Delta_{t-1}^{-1}a_{t-1} > 0,$$

for $\Delta_{t-1}\in(0,1]$. The recursive step for q_t is given by

$$q_t = \frac{\Delta_t a_{1|0}}{a_t(1-\Delta_t^2)+\Delta_t^2 a_{1|0}} > 0,$$

for $\Delta_t \in (0,1]$. This proves that $q_t \in \mathbb{R}_+$ and $a_t \in \mathbb{R}_+$ for all $t \geq 1$.

Theorem 1. *Under the framework of Model 2, the inverse state-space process $(\Theta_t^{-1})_{t\geq 1}$ is stationary in both its mean and variance..*

Proof of Theorem 1. By Proposition 2, we have the mean stationarity of $(\Theta_t^{-1})_{t\geq 1}$. It remains to prove that this model gives a finite variance process. For this, we require that $a_t > 1$ and $a_{t|t-1} > 1$ for all $t \geq 1$. The former follows from the latter because $a_t \geq a_{t|t-1}$. So it remains to prove the latter. We initialize $a_{1|0} > 1$. This implies $a_1 \geq a_{1|0} > 1$. From this we can see that for $t = 2$

$$a_{2|1} = q_1\Delta_1^{-1}a_1 = \frac{a_{1|0}}{a_1-\Delta_1^2 a_1+\Delta_1^2 a_{1|0}}a_1 = \frac{a_{1|0}}{a_1-\Delta_1^2(a_1-a_{1|0})}a_1 > a_{1|0} > 1.$$

This implies $a_2 \geq a_{2|1} > a_{1|0} > 1$. Recursive iteration for $t \geq 3$ proves the claim. □

1708945x, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjs.70024 by Ewha Womans University Library, Wiley Online Library on [21/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4.3 Model with an increasing variance process

Proposition 1 proves that the Smith–Miller model exhibits an increasing variance behaviour in the inverse state-space dynamics. Consequently, Model 1 is capable of accommodating such a behaviour, as the Smith–Miller model is a special case of Model 1. That is, the Smith–Miller model corresponds to Model 1 under the following parameter settings, for $t \geq 1$,

$$q_t = \frac{\gamma(a_t - 1) + 1}{a_t} \qquad \text{and} \qquad p_t = 0.$$

We extend this result by showing that under suitable exogenous sequences in Model 1, the variance not only increases but even diverges.

Lemma 3. *Consider model 1 with further conditions $p_t = 0$ and $v_t = 1$ for all $t \geq 1$, and dispersion parameter $\psi = 1$. Moreover, assume $0 < q_t < \delta < 1$ for all $t \geq 1$. Then, we have*

$$\liminf_{t\to\infty} \operatorname{var}[\Theta_t^{-1}] = \infty. \tag{22}$$

Proof of Lemma 3. Assume that there exists $N > 0$ such that $a_t > 1$ for all $t \geq N$. If not, equation (22) holds because the variance can only be finite if $a_t > 1$. That is, in this case we have $\operatorname{var}[\Theta_t^{-1}] < \infty$ for all $t \geq N$. Then we come back to Lemma 1. Setting $p_t = 0$, $t \geq 1$, we have from (19)

$$\operatorname{var}\left[\Theta_{t+1}^{-1}\right] = \frac{q_t a_t - q_t}{q_t a_t - 1} \operatorname{var}[\Theta_t^{-1}] + (1 - q_t)\frac{1}{q_t a_t - 1}, \tag{23}$$

for all $t \geq N$. Consider the sequence $(a_t)_{t\geq 1}$

$$a_{t+1} = a_{t+1|t} + v_{t+1}/\psi = q_t a_t + 1 \leq \delta a_t + 1 \leq \delta^t a_{1|0} + \sum_{k=0}^{t} \delta^k \leq a_{1|0} + \frac{\delta}{1-\delta} < \infty,$$

for all $t \geq N$. This implies that the second term on the right-hand side of (23) is bounded uniformly from below by a strictly positive constant. Moreover, the first ratio on the right-hand side of (23) is bounded from below by 1, and therefore (23) implies that a sequence $\left(\operatorname{var}[\Theta_t^{-1}]\right)_{t\geq 1}$ is diverging for $t \to \infty$. This completes the proof. □

4.4 Model with a decreasing variance process

Next, we present a setting of Model 1 that has decreasing variances of the inverse state-space process. In the extreme case, these variances converge to zero, see Corollary 1 below. For this, we consider model 1 under the constraint

$$p_t + q_t = 1 \qquad \text{for all} \quad t \geq 1, \tag{24}$$

with $p_t \in [0, 1)$ for $t \geq 1$.

Lemma 4. *Consider model 1 with initialization $a_{1|0} > 1$ and with constraint (24). Then,*

$$\operatorname{var}[\Theta_{t+1}^{-1}] \leq \operatorname{var}[\Theta_t^{-1}], \qquad \text{for } t \geq 1,$$

where the equality holds if and only if $p_t = 0$.

Proof of Lemma 4. First we note that $\Delta_t \in (0, 1]$ is deterministic, and $a_t, a_{t|t-1} > 0$ for all $t \geq 1$. Formula (18) then implies

$$\operatorname{var}[\mathrm{E}[\Theta_{t+1}^{-1} \mid Y_{1:t}]] = \Delta_t^2 \operatorname{var}[\mathrm{E}[\Theta_t^{-1} \mid Y_{1:t}]] \leq \operatorname{var}[\mathrm{E}[\Theta_t^{-1} \mid Y_{1:t}]],$$

and we have an equality if and only if $\Delta_t = 1$, which is equivalent to $p_t = 0$. Next, observe

$$\begin{aligned}
\mathrm{E}\Big[\operatorname{var}[\Theta_{t+1}^{-1} \mid Y_{1:t}]\Big] &= \mathrm{E}\left[\frac{b_{t|t-1}^2}{a_{t|t-1}^2(a_{t|t-1} - 1)}\right] = \mathrm{E}\left[\frac{(p_t a_t + q_t b_t)^2}{((p_t + q_t)a_t)^2((p_t + q_t)a_t - 1)}\right] \\
&= \frac{\mathrm{E}\Big[((1 - q_t)a_t + q_t b_t)^2\Big]}{a_t^2(a_t - 1)} \leq \frac{\mathrm{E}[b_t^2]}{a_t^2(a_t - 1)} = \mathrm{E}\Big[\operatorname{var}[\Theta_t^{-1} \mid Y_{1:t}]\Big].
\end{aligned}$$

1708945x, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjs.70024 by Ewha Womans University Library, Wiley Online Library on [21/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The inequality follows from mean stationarity

$$a_t = a_t \mathrm{E}\left[\Theta_t^{-1}\right] = a_t \mathrm{E}\left[\mathrm{E}\left[\Theta_t^{-1} \middle| Y_{1:t}\right]\right] = a_t \mathrm{E}\left[\frac{b_t}{a_t}\right] = \mathrm{E}[b_t],$$

which together with Jensen's inequality implies

$$\begin{aligned}
\mathrm{E}\left[\left((1-q_t)a_t + q_t b_t\right)^2\right] &= (1-q_t)^2 a_t^2 + 2q_t(1-q_t)a_t^2 + \mathrm{E}\left[q_t^2 b_t^2\right] \\
&= (1-q_t)^2 \mathrm{E}[b_t]^2 + 2q_t(1-q_t)\mathrm{E}[b_t]^2 + \mathrm{E}\left[q_t^2 b_t^2\right] \\
&\leq (1-q_t)^2 \mathrm{E}\left[b_t^2\right] + 2q_t(1-q_t)\mathrm{E}\left[b_t^2\right] + \mathrm{E}\left[q_t^2 b_t^2\right] \\
&= \mathrm{E}\left[b_t^2\right],
\end{aligned}$$

and we have a strict inequality whenever $q_t < 1$, which is equivalent to $p_t > 0$ under $p_t + q_t = 1$. This proves the claim. □

In the next result we prove that if, under the assumptions of Lemma 4, the sequence $(q_t)_{t\geq 1}$ is bounded away from 0 and 1, the inverse state-space process has an asymptotically vanishing variance, i.e., asymptotically the randomness from the state-spaces is zero.

Corollary 1. *Consider model 1 with initialization $a_{1|0} > 1$ and with constraint (24). Moreover, assume $\delta < q_t^2 < 1 - \delta$ for some $\delta > 0$ and all $t \geq 1$, and $\sum_{t\geq 1} v_t = \infty$. Then, we have*

$$\lim_{t\to\infty} \mathrm{var}[\Theta_t^{-1}] = 0.$$

Proof of Corollary 1. From Lemma 1 and under constraint (24), we have

$$\mathrm{var}[\Theta_{t+1}^{-1}] = q_t^2 \,\mathrm{var}[\Theta_t^{-1}] + \left(1 - q_t^2\right)\frac{1}{a_t - 1} \leq (1-\delta)\,\mathrm{var}[\Theta_t^{-1}] + (1-\delta)\frac{1}{a_t - 1},$$

for $t \geq 1$. Choosing $t + s > t \geq 1$, we have, using the monotonicity established in Lemma 4

$$\begin{aligned}
\mathrm{var}[\Theta_{t+s}^{-1}] &\leq (1-\delta)^s \,\mathrm{var}[\Theta_t^{-1}] + \sum_{k=1}^{s} (1-\delta)^k \frac{1}{a_{t+s-k} - 1} \\
&\leq (1-\delta)^s \,\mathrm{var}[\Theta_1^{-1}] + \left(\frac{1-\delta}{\delta}\right) \max_{1\leq k\leq s} \frac{1}{a_{t+s-k} - 1}.
\end{aligned}$$

Using (14) and (16), we have, for $t \geq 2$

$$a_t = a_{t|t-1} + v_t/\psi = (p_{t-1} + q_{t-1})a_{t-1} + v_t/\psi = a_{t-1} + v_t/\psi = a_{1|0} + \psi^{-1} \sum_{s=1}^{t} v_s.$$

Under our assumptions, this implies monotonicity of $(a_t)_{t\geq 1}$ with $\lim_{t\to\infty} a_t = \infty$. Plugging this into the previous inequality, we obtain the upper bound

$$\mathrm{var}[\Theta_{t+s}^{-1}] \leq (1-\delta)^s \,\frac{1}{a_{1|0} - 1} + \left(\frac{1-\delta}{\delta}\right)\frac{1}{a_t - 1}.$$

Using $\lim_{t\to\infty} a_t = \infty$, we can find, for any $\varepsilon > 0$, an index $t_0 \geq 1$ such that the second term satisfies

$$\left(\frac{1-\delta}{\delta}\right)\frac{1}{a_{t_0} - 1} < \varepsilon/2.$$

Moreover, there exists $s_0 > 0$ such that for all $s \geq s_0$

$$(1-\delta)^s \,\frac{1}{a_{1|0} - 1} < \varepsilon/2.$$

1708945x, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjs.70024 by Ewha Womans University Library, Wiley Online Library on [21/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

But this implies that for all $u \geq s_0 + t_0$ we have

$$\text{var}[\Theta_u^{-1}] < \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, the claim follows. □

Remark 3. For $p_t \equiv 0$ and $q_t \equiv 1$, there is no state-space update from t to $t+1$ meaning that

$$\Theta_t | Y_{1:t} \stackrel{\text{d}}{=} \Theta_{t+1} | Y_{1:t}.$$

In this case, Model 1 becomes a static random effects model, which complies with the Bühlmann and Straub (1970) assumptions. In particular, it has one static latent factor $\Theta_t \equiv \Theta \sim \gamma(a_{1|0}, a_{1|0})$; see (Bühlmann and Gisler, 2005, page 44).

From the previous results, we conclude that Model 1 allows for a wide range of variance behaviours for the inverse state-space dynamics $(\Theta_t^{-1})_{t\geq 1}$. The specific variance behaviour should be determined by the data through model fitting and model selection. This is discussed next.

5 Model fitting, forecasting, and evolutionary credibility

5.1 Evolutionary credibility and credibility formulas

For experience rating, one is mainly interested in forecasting the claim in the next period Y_{t+1}, given the past information $Y_{1:t}$. Under model 1, we can compute the conditional expectation of the future claim, given the past observations, by applying the tower property for conditional expectations. That is,

$$\mathrm{E}[\, Y_{t+1} | Y_{1:t}] = \mathrm{E}[\, \mathrm{E}[\, Y_{t+1} | Y_{1:t}, \Theta_{1:t+1}] | Y_{1:t}] = v_{t+1}\mu_{t+1}\, \mathrm{E}\Big[\, \Theta_{t+1}^{-1} \Big| Y_{1:t}\Big].$$

Using (15) and mean stationarity, we arrive at

$$\mathrm{E}[\, Y_{t+1} | Y_{1:t}] = v_{t+1}\mu_{t+1} \left(\Delta_t \mathrm{E}[\Theta_t^{-1} \mid Y_{1:t}] + (1-\Delta_t)\mathrm{E}[\Theta_{t+1}^{-1}] \right). \tag{25}$$

Thus, the posterior prediction is a credibility-weighted average between the observation-driven estimate $\mathrm{E}[\Theta_t^{-1} \mid Y_{1:t}]$ and the prior estimate $\mathrm{E}[\Theta_{t+1}^{-1}] = \mathrm{E}[\Theta_1^{-1}] = 1$, with credibility weight, see (17),

$$\Delta_t = \frac{q_t}{q_t + p_t} \in (0, 1].$$

This recursive formula (25), along with Bayesian filtering in equation (14), provides a basis for deriving the evolutionary credibility structure (Bühlmann and Gisler, 2005, Chapter 9). Specifically, starting from (11), we obtain the posterior mean

$$\begin{aligned}
\mathrm{E}[\Theta_t^{-1} \mid Y_{1:t}] = \frac{b_t}{a_t} &= \frac{b_{t|t-1} + Y_t/(\mu_t\psi)}{a_{t|t-1} + v_t/\psi} \\
&= \begin{cases} \frac{v_t/\psi}{a_{t|t-1}+v_t/\psi} \frac{1}{\mu_t} \frac{Y_t}{v_t} + \frac{a_{t|t-1}}{a_{t|t-1}+v_t/\psi} \frac{b_{t|t-1}}{a_{t|t-1}}, & \text{for } v_t > 0; \\ \frac{b_{t|t-1}}{a_{t|t-1}}, & \text{for } v_t = 0. \end{cases}
\end{aligned}$$

To simplify the notation, we define

$$\frac{Y_t}{v_t} = 0, \qquad \text{for} \quad Y_t = 0 \quad \text{and} \quad v_t = 0$$

and

$$\frac{1}{\mu_t} \frac{\mathrm{E}[\, Y_t | Y_{1:t-1}]}{v_t} = \mathrm{E}\Big[\, \Theta_t^{-1} \Big| Y_{1:t-1}\Big], \qquad \text{for} \quad v_t = 0.$$

With this definition, the recursion can be compactly expressed as

$$\mathrm{E}[\Theta_t^{-1} \mid Y_{1:t}] = \frac{v_t/\psi}{a_{t|t-1} + v_t/\psi} \frac{1}{\mu_t} \frac{Y_t}{v_t} + \frac{a_{t|t-1}}{a_{t|t-1} + v_t/\psi}\, \mathrm{E}[\Theta_t^{-1} \mid Y_{1:t-1}].$$

1708945x, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjs.70024 by Ewha Womans University Library, Wiley Online Library on [21/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Combining this with (12) and (25), we obtain the evolutionary credibility formula for the normalized observations $(Y_t/(\upsilon_t\mu_t))_{t\geq 1}$.

Proposition 3. *Assume Model 1 holds. Define the weights z_t by*

$$z_t = \frac{\upsilon_t/\psi}{a_{t|t-1} + \upsilon_t/\psi} \in [0,1).$$

For $t \geq 1$, the predictive conditional mean of the future claim is given by

$$\begin{aligned}\mathrm{E}[Y_{t+1}|Y_{1:t}] &= \upsilon_{t+1}\mu_{t+1}\,\mathrm{E}\left[\Theta_{t+1}^{-1}\middle|Y_{1:t}\right] \\ &= \upsilon_{t+1}\mu_{t+1}\left(c_{1,t}\frac{1}{\mu_t}\frac{Y_t}{\upsilon_t} + c_{2,t}\frac{1}{\mu_t}\frac{\mathrm{E}[Y_t|Y_{1:t-1}]}{\upsilon_t} + c_{3,t}\right),\end{aligned}$$

where the weights $c_{1,t}, c_{2,t}, c_{3,t} \geq 0$ satisfy $c_{1,t} + c_{2,t} + c_{3,t} = 1$ and are given by

$$c_{1,t} = \Delta_t z_t, \quad c_{2,t} = \Delta_t(1-z_t), \quad \text{and} \quad c_{3,t} = 1 - \Delta_t.$$

By recursively expanding the previous result, we arrive at the following corollary.

Corollary 2. *Under model 1, the predictive conditional mean can be represented by*

$$\mathrm{E}[Y_{t+1}|Y_{1:t}] = \upsilon_{t+1}\mu_{t+1}\left(\sum_{s=1}^{t}\left[\prod_{k=s+1}^{t} c_{2,k}\right]\left(c_{1,s}\left[\frac{1}{\mu_s}\frac{Y_s}{\upsilon_s}\right] + c_{3,s}\cdot 1\right) + \prod_{s=1}^{t} c_{2,s}\right),$$

for $t \geq 0$, *where empty products and sums represent 1 and 0, respectively.*

By setting

$$\Delta_0 = 0 \quad \text{and} \quad \frac{1}{\mu_0}\frac{Y_0}{\upsilon_0} = 0,$$

the predictive conditional mean in Corollary 2 can be simplified as

$$\begin{aligned}\mathrm{E}[Y_{t+1}|Y_{1:t}] &= \upsilon_{t+1}\mu_{t+1}\left(\sum_{s=0}^{t}\left[\prod_{k=s+1}^{t} c_{2,k}\right]\left(c_{1,s}\left[\frac{1}{\mu_s}\frac{Y_s}{\upsilon_s}\right] + c_{3,s}\right)\right) \\ &= \upsilon_{t+1}\mu_{t+1}\left(\omega_{0,t} + \sum_{s=1}^{t}\omega_{s,t}\left[\frac{1}{\mu_s}\frac{Y_s}{\upsilon_s}\right]\right), \qquad (26)\end{aligned}$$

where the weights are defined as

$$\omega_{s,t} = \left(\prod_{k=s+1}^{t} c_{2,k}\right)c_{1,s}, \quad s \in \{1,\dots,t\},$$

and

$$\omega_{0,t} = \prod_{k=1}^{t} c_{2,k} + \sum_{s=1}^{t}\left(\prod_{k=s+1}^{t} c_{2,k}\right)c_{3,s}.$$

Clearly, we have

$$\omega_{0,t}, \omega_{1,t}, \dots, \omega_{t,t} \geq 0 \quad \text{and} \quad \sum_{s=0}^{t}\omega_{s,t} = 1.$$

Thus, we obtain a credibility-weighted average between the observations $Y_{1:t}$ and the prior mean $\mathrm{E}[\Theta_{t+1}^{-1}] = 1$. Here, we call $\omega_{1,t}, \dots, \omega_{t,t}$ the credibility weights corresponding to the observations $Y_1, \dots, Y_t$. From an insurance point of view, there are two main points that make this Generalized Smith–Miller model of interest. First, the different variance behaviours will result in different risk capital requirements because uncertainty can be increasing or decreasing over time depending on the choices of the exogenous parameters. Second,

17089458x, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjs.70024 by Ewha Womans University Library, Wiley Online Library on [21/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

these different choices of the exogenous parameters directly enter the above predictive formula, resulting in different weighting schemes of past claims. In particular, specific exogenous parameter choices may lead to an exponentially decaying seniority weighting in past claims, which seems a reasonable model property in insurance pricing. Basically, it means that past claims enter through an exponentially fading memory property. For seniority weighting, see Pinquet et al. (2001).

5.2 Likelihood function

The joint density of the observations $Y_{1:t}$ under Model 1 is given by the recursive formula

$$f(y_{1:t}) = f(y_t|y_{1:t-1})f(y_{1:t-1}) = \prod_{s=1}^{t} f(y_s|y_{1:s-1}),$$

where $f(y_s|y_{1:s-1})$ for $2 \le s \le t$ represents the conditional density of Y_s in y_s, for given $Y_{1:s-1} = y_{1:s-1}$, and $f(y_1) = f(y_1|y_{1:0})$ represents the density of Y_1 in y_1. We have the following result.

Lemma 5. *Under model 1, for $s \ge 1$, the following holds.*

i. The conditional density of Y_s in $y_s > 0$, given observations $Y_{1:s-1} = y_{1:s-1}$, is given by

$$f(y_s \mid y_{1:s-1}) = \frac{\gamma(v_s/\psi + a_{s|s-1} + 1)}{\gamma(v_s/\psi)\gamma(a_{s|s-1}+1)} \left(\frac{\frac{y_s}{\mu_s\psi}}{\frac{y_s}{\mu_s\psi} + b_{s|s-1}}\right)^{v_s/\psi} \left(\frac{b_{s|s-1}}{\frac{y_s}{\mu_s\psi} + b_{s|s-1}}\right)^{a_{s|s-1}+1} y_s^{-1}.$$

ii. Define the random variable

$$X_s = \frac{Y_s}{\mu_s\psi b_s}.$$

The conditional density of X_s in $x_s > 0$, given observations $Y_{1:s-1} = y_{1:s-1}$, is given by

$$f(x_s \mid y_{1:s-1}) = \frac{\gamma(v_s/\psi + a_s + 1)}{\gamma(v_s/\psi)\gamma(a_s+1)} x_s^{v_s/\psi - 1}(1+x_s)^{v_s/\psi + a_s + 1}.$$

This is a Beta-prime distribution with parameters $v_s/\psi > 0$ and $a_s + 1 > 0$, also called a Pearson type VI distribution (Johnson et al., 1995).

Proof of Lemma 5. Using (2) and (12), we have for $s \ge 1$ and $y_s > 0$

$$\begin{aligned}
f(y_s \mid y_{1:s-1}) &= \int_{\Theta_s} f(y_s \mid \theta_s) f(\theta_s \mid y_{1:s-1})\, d\theta_s \\
&= \frac{\left(\frac{1}{\mu_s\psi}\right)^{\frac{v_s}{\psi}}}{\gamma\left(\frac{v_s}{\psi}\right)} y_s^{\frac{v_s}{\psi}-1} \frac{\left(b_{s|s-1}\right)^{a_{s|s-1}+1}}{\gamma(a_{s|s-1}+1)} \int_{\Theta_s} \theta_s^{\frac{v_s}{\psi}+a_{s|s-1}+1-1} \exp\left\{-\theta_s\left(\frac{y_s}{\mu_s\psi} + b_{s|s-1}\right)\right\} d\theta_s \\
&= \frac{\left(\frac{1}{\mu_s\psi}\right)^{\frac{v_s}{\psi}}}{\gamma\left(\frac{v_s}{\psi}\right)} y_s^{\frac{v_s}{\psi}-1} \frac{\left(b_{s|s-1}\right)^{a_{s|s-1}+1}}{\gamma(a_{s|s-1}+1)} \frac{\gamma(\frac{v_s}{\psi} + a_{s|s-1}+1)}{\left(\frac{y_s}{\mu_s\psi} + b_{s|s-1}\right)^{\frac{v_s}{\psi}+a_{s|s-1}+1}} \\
&= \frac{\gamma(\frac{v_s}{\psi} + a_{s|s-1}+1)}{\gamma(\frac{v_s}{\psi})\gamma(a_{s|s-1}+1)} \left(\frac{\frac{y_s}{\mu_s\psi}}{\frac{y_s}{\mu_s\psi} + b_{s|s-1}}\right)^{\frac{v_s}{\psi}} \left(\frac{b_{s|s-1}}{\frac{y_s}{\mu_s\psi} + b_{s|s-1}}\right)^{a_{s|s-1}+1} y_s^{-1}.
\end{aligned}$$

This completes the proof of part (i).

The proof of the second part follows by a change of variables for $X_s = Y_s/(\mu_s\psi b_s)$. □

1708945x, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjs.70024 by Ewha Womans University Library, Wiley Online Library on [21/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Note that

$$\lim_{y_s\to\infty}\frac{f(y_s \mid y_{1:s-1})}{y_s^{-(a_{s|s-1}+2)}} = \frac{\gamma(v_s/\psi + a_{s|s-1}+1)}{\gamma(v_s/\psi)\gamma(a_{s|s-1}+1)}\left(\mu_s\psi b_{s|s-1}\right)^{a_{s|s-1}+1} \in (0,\infty).$$

This implies that the conditional distribution of Y_s, given $Y_{1:s-1}$, is regularly varying at infinity with tail index $-(a_{s|s-1}+2)$. Applying Karamata's theorem (Embrechts et al., 2013) to the corresponding survival function shows that the conditional survival function of Y_s, given $Y_{1:s-1}$, is regularly varying at infinity with tail index $-(a_{s|s-1}+1)$. In this regard, the proposed state-space modelling framework can be effectively utilized to analyze claims with a heavy tail, although they originate from the combinations of light-tailed gamma random variables. The asymptotic properties also provide conditions for the finiteness of the conditional moments. In particular, if $a_{s|s-1} > 0$, then $\mathrm{E}[Y_s|Y_{1:s-1}]$ is finite, and if $a_{s|s-1} > 1$, then $\mathrm{var}[Y_s|Y_{1:s-1}]$ is finite. Naturally, these observations are consistent with the earlier conclusions in Lemma 1 regarding the existence of the conditional variance of the inverse state-space variable Θ_s^{-1}, from which the present (conditional) likelihood results are derived.

In case of a unit exposure $v_t \equiv 1$ and a unit dispersion $\psi = 1$, the gamma distribution (2) turns into the simpler exponential distribution. As a consequence, we get the conditional distribution

$$f(y_s \mid y_{1:s-1}) = \frac{a_{s|s-1}+1}{\mu_s b_{s|s-1}}\left(1+\frac{y_s}{\mu_s b_{s|s-1}}\right)^{-(a_{s|s-1}+2)}.$$

This is a Pareto Type II distribution, also called the Lomax distribution (Lomax, 1954), with shape parameter $a_{s|s-1}+1 > 1$ and scale parameter $\mu_s b_{s|s-1} > 0$.

Appendix A presents a simulation study, and in the next section we present a real data analysis. For this, we equip the claims $(Y_{i,t})_{t\geq 1}$ with a second index i labelling the different policyholders $1 \leq i \leq M$.

6 Real-data analysis

To assess the applicability of the proposed framework, we analyze a US-based longitudinal outpatient visit dataset for the years 2019–2022 (on a half-year time scale). These data are parts of the database called the Medical Expenditure Panel Survey (MEPS). The original dataset, which contains 30,079 records for 4928 patients over the years 2019–2022, is a combination of the following four tables in the MEPS database: HC-213F, HC-220F, HC-229F, and HC-239F. As displayed in Table 1, it contains various fields such as the identifier for a patient in the panel (`DUPERSID`), year and month information for the outpatient visit (`OPDATEYR` and `OPDATEMM`), categorical covariates that explain the characteristics of the outpatient visit (for example, `SEEDOC_M18` and `LABTEST_M18`; see the codebook to check the comprehensive list of the covariates and their descriptions), and the total medical expenses charged for the visit (`OPTC[YR]X`).

6.1 Data preprocessing with a working model

While the dataset contains rich information with repeated measurements, it required some preprocessing to be analyzed under the proposed framework. In the Generalized Smith–Miller model, we implicitly assume that the state-space variables $(\Theta_t)_{t\geq 1}$ affect each period of the same length (for example, a year or a month) whereas the original dataset was not recorded in that fashion. In other words, the observed value of `OPTC[YR]X` in each record from the original dataset corresponds to an individual payment $Z_{t[j]}$, as described in equation (7), not to Y_t in (2) for a time period.

Table 1: Sample rows from the original dataset.

DUPERSID	OPDATEYR	OPDATEMM	SEEDOC_M18	LABTEST_M18	⋯	OPTC[YR]X
2460002101	2021	8	Y	Y	⋯	6407.5
2460002101	2022	10	Y	Y	⋯	5344
⋮	⋮	⋮	⋮	⋮	⋱	⋮
2460004102	2021	3	N	N	⋯	83.01
2460004102	2021	10	N	N	⋯	100.01

1708945x, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjs.70024 by Ewha Womans University Library, Wiley Online Library on [21/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2: Description of the covariates in the working model.

Variables	Description	Response	Proportions
`Doctor_Type`	Type of medical professional for the visit	GP	7.00 %
		Specialist	39.27 %
		Non-doctor	53.73 %
`Care_Category`	Main reason of the outpatient visit	Diagnosis or Treatment	51.67 %
		Others	48.33 %
`Special_Cond`	Indicator for an existing special condition for the visit	Yes	83.62 %
		No	16.38 %
`Surgery`	Whether the visit involved a surgery	Yes	8.68 %
		No	91.32 %
`Prescription`	Whether any medicine prescribed for the visit	Yes	8.33 %
		No	91.67 %
`Telehealth`	Whether the visit was a telehealth event	Yes	7.61 %
		No	40.35 %
		Unknown	52.04 %

Therefore, the observed records for a person i (identified by `DUPERSID`, here) needed to be aggregated within prespecified periods $t \geq 1$ to give us $(Y_{i,t}, v_{i,t}, \mu_{i,t})_{i,t}$, where $Y_{i,t}$ is the sum of `OPTC[YR]X` for person i in period t, $v_{i,t}$ is the total number of the outpatient visits for person i in period t, and $\mu_{i,t} = \mathbb{E}[Z_{t[j]}]$ is the expectation of `OPTC[YR]X` per outpatient visit of person i in period t. While the aggregations of $Y_{i,t}$ and $v_{i,t}$ are straightforward and do not involve any uncertainty, $\mu_{i,t}$ is unknown at this stage and so we needed a working model that estimates the marginal mean severity $\mu_{i,t}$ based on the observed values of the response variables $Z_{i,t[j]}$ (which is `OPTC[YR]X` in this case) and the corresponding covariates in the original dataset.

For this purpose, we exploited a preliminary gamma generalized linear model (GLM) assuming

$$Z_{i,t[j]} \sim \gamma\left(\frac{1}{\psi}, \frac{1}{\widetilde{\mu}_{i,t[j]}\psi}\right), \qquad \text{with regression mean: } \log(\widetilde{\mu}_{i,t[j]}) = \mathbf{x}_{i,t[j]}\boldsymbol{\beta},$$

where $\mathbf{x}_{i,t[j]}$ are preprocessed covariates from the original dataset as described in Table 2, and GLM parameter $\boldsymbol{\beta}$. When the covariates were preprocessed for the working model, missing values in `Care_Category` were included in the category 'Others', and for the variables `Special_Cond`, Surgery, and Prescription these values were classified as 'No'. The proportion of such cases was negligible (less than 1%). In the case of `Telehealth`, however, we treated the missing values as a separate category, because this covariate was not recorded in 2019 and hence all the values are missing for this year. Moreover, about 25% of records from years 2020–2022 had missing values for this variable.

The estimation results of the working model are summarized in Table 3, which also provides some intuitive interpretations. For example, it is expected that an outpatient visit will incur a higher charge if it involves a meeting with a specialist, rather than a general or family doctor. It is also natural to expect that outpatient visits with surgeries will incur higher charges than those without surgeries. Lastly, it is also shown that a telehealth outpatient visit incurs a lower charge than an in-person outpatient visit.

Note that it is possible that $\mathbf{x}_{i,t[j]}$ may not be identical for all $j = 1, \ldots, v_{i,t}$, because in a given time period a patient may have multiple visits for different reasons. Therefore, we summarized the estimates from the working model $\widetilde{\mu}_{i,t[j]}$ to obtain a reasonable estimate of $\mu_{i,t}$ as $\widehat{\mu}_{i,t} = 1/v_{i,t} \sum_{j=1}^{v_{i,t}} \widetilde{\mu}_{i,t[j]}$. Sample rows from the resulting dataset in the form of $(Y_{i,t}, v_{i,t}, \mu_{i,t})_{i,t}$ are displayed in Table 4. Note that we summarized the observations per

1708945x, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjs.70024 by Ewha Womans University Library, Wiley Online Library on [21/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 3: Summary of the estimated regression coefficients for the working severity model.

	Estimate	*P*-value
(Intercept)	7.4351	0.0000
`Doctor_Type`: Non-doctor	0.1304	0.0338
`Doctor_Type`: Specialist	0.5515	0.0000
`Care_Category`: Others	−0.1390	0.0000
`Special_Cond`: Yes	0.1912	0.0000
`Surgery`: Yes	1.6888	0.0000
`Prescription`: Yes	−0.1080	0.0531
`Telehealth`: Unknown	−0.1501	0.0000
`Telehealth`: Yes	−0.8352	0.0000

Table 4: Sample rows from the preprocessed dataset.

ID	HY	$Y_{i,t}$	$v_{i,t}$	$\widehat{\mu}_{i,t}$
2460002101	2019H1	1209.00	1	1750.443
2460002101	2019H2	13712.17	4	2011.496
2460002101	2020H1	0	0	–
2460002101	2020H2	0	0	–
⋮	⋮	⋮	⋮	⋮

half-year intervals so that the preprocessed dataset has $39,424 = 4928 \times 8$ records, from 4928 distinct patients and $t = 8$ observational periods (from 2019H1 to 2022H2).

We conclude this subsection with two remarks. Firstly, it is not our primary concern to find either the best working model or the best set of covariates to estimate the marginal severity, `OPTC[YR]X`. Possibly, one can conduct a more sophisticated analysis to find the best set of engineered features and/or try different predictive models (including but not limited to random forests or neural networks). However, we believe that the current working model is reasonable enough to quantify the underlying impacts of the available covariate information on the response variables, with some intuitive explanations. Note also that our main goal is to evaluate the usefulness of the proposed Generalized Smith–Miller model framework in comparison to the existing models that could describe the serial correlations among the repeated observations, given $\widehat{\mu}_{i,t}$ from some reasonably good working model.

6.2 Analysis of the summarized dataset with the proposed model and some benchmarks

For comparison purposes, we study six different models. First, we have two sets of models, so-called homogeneous ones and heterogeneous ones. For the homogeneous models we do not consider any covariates, and we set the homogeneous mean parameter μ to the empirical weighted average of $(Y_{i,t})$, that is,

$$\widehat{\mu} = \frac{\sum_{i=1}^{M}\sum_{t=1}^{T} Y_{i,t}}{\sum_{i=1}^{M}\sum_{t=1}^{T} v_{i,t}}.$$

For the heterogeneous models, we select the estimated means $\widehat{\mu}_{i,t}$ as outlined in Table 3. For each of these two different cases (homogeneous and heterogeneous), we select three different models.

- Independent: We assume that $Y_{i,t}\big|Y_{i,1:t-1}$ follows a gamma distribution as in equation (2) with $\theta_{i,t} \equiv 1$ for $1 \le t \le 8$ and $1 \le i \le M$. That is, there is no dependence across time.
- Bühlmann model: We assume that $Y_{i,t}\big|Y_{i,1:t-1}$ follows a gamma distribution as in equation (2) with $\theta_{i,t} = \theta_i$ for $1 \le t \le 8$ and $1 \le i \le M$. Thus, we have a static latent risk factor, and past observations do

1708945x, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjs.70024 by Ewha Womans University Library, Wiley Online Library on [21/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

not receive a seniority weighting, but they are exchangeable for prediction. Note that this is a special case of the next model with $\Delta = 1$.

- Observation-driven state-space model (SSM): We assume that $Y_{i,t}\big|Y_{i,1:t-1}$ follows a gamma distribution as in equation (2), allowing for an evolution of $\theta_{i,1:8}$ under model 1, assuming (20) holds.

Recall that all of these models (implicitly or explicitly) assume a stationary variance for $(\Theta_{i,t}^{-1})_{t\geq 1}$ by satisfying (20). So, if we assume independence between the claims of different policyholders, then the joint log-likelihood of the constant variance SSM in Model 2 can be expressed as a function of $a_{1|0}$ and ψ (note that a constant variance model is enough for our real-data analysis, as its time horizon is too short, 8 periods, to observe drastic changes in the variances):

$$\ell(a_{1|0},\psi) = \sum_{i=1}^{M}\sum_{s=1}^{t} \log f\big(y_{i,s}|y_{i,1:s-1}\big), \tag{27}$$

where the policyholder-specific weights $v_{i,s} \in \mathbb{N}_0$ and means $\mu_{i,s} > 0$ are given exogenously for $i \in \{1,\dots,M\}$ and $s \in \{1,\dots,t\}$, along with the exogenous sequence $(\Delta_t)_{t\geq 1}$.

Note also that the exogenous sequence $(q_t)_{t\geq 1}$ is implied by $(\Delta_t)_{t\geq 1}$ via the constant-variance condition in equation (21). The individual terms in the log-likelihood are given by Lemma 5, allowing for maximum likelihood estimation (MLE) based on the observed data.

While Model 2 assumes the given exogenous sequence $(\Delta_t)_{t\geq 1}$, selecting an appropriate sequence may be challenging in practice. To mitigate this difficulty, we assume

$$\Delta_t = \Delta$$

for some constant $\Delta \in (0,1]$. This form of constant credibility weighting has been widely adopted in insurance applications; see, for instance, Gerber and Jones (1975a, 1975b), and Sundt (1987). Under this simplification, the log-likelihood in equation (27) becomes fully specified by the policyholder-dependent exogenous sequences

$$v_{i,1:t} \quad \text{and} \quad \mu_{i,1:t}, \quad \text{for } i \in \{1,\dots,M\},$$

together with the model parameters $a_{1|0}$, ψ, and the additional parameter Δ:

$$\ell(a_{1|0},\psi,\Delta) = \sum_{i=1}^{M}\sum_{s=1}^{t} \log f\big(y_{i,s}|y_{i,1:s-1}\big). \tag{28}$$

Table 5 summarizes the estimation results of $a_{1|0}$, ψ, and Δ using the MLE based on the log-likelihood function (28) and based on the aggregated version of the actual data as described above. The homogeneous/heterogeneous independent models are excluded in Table 5, as these models do not require estimation of $a_{1|0}$, ψ, and Δ (by assuming θ is constantly equal to 1). It is shown that the estimated values of Δ under the homogeneous/heterogeneous SSMs are around 0.3, which implies a rather fast decay of the impact of

Table 5: Summary of the parameter estimation on the real dataset (with standard errors in parentheses).

	$\widehat{a}_{1\|0}$	$\widehat{\psi}$	$\widehat{\Delta}$
Homogeneous Bühlmann	0.8257	2.0394	1.0000
	(0.0291)	(0.0304)	
Homogeneous SSM	0.4166	0.8746	0.3262
	(0.0110)	(0.0274)	(0.0302)
Heterogeneous Bühlmann	1.4543	1.7334	1.0000
	(0.0575)	(0.0257)	
Heterogeneous SSM	0.7150	0.8257	0.2825
	(0.0202)	(0.0238)	(0.0303)

1708945x, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjs.70024 by Ewha Womans University Library, Wiley Online Library on [21/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

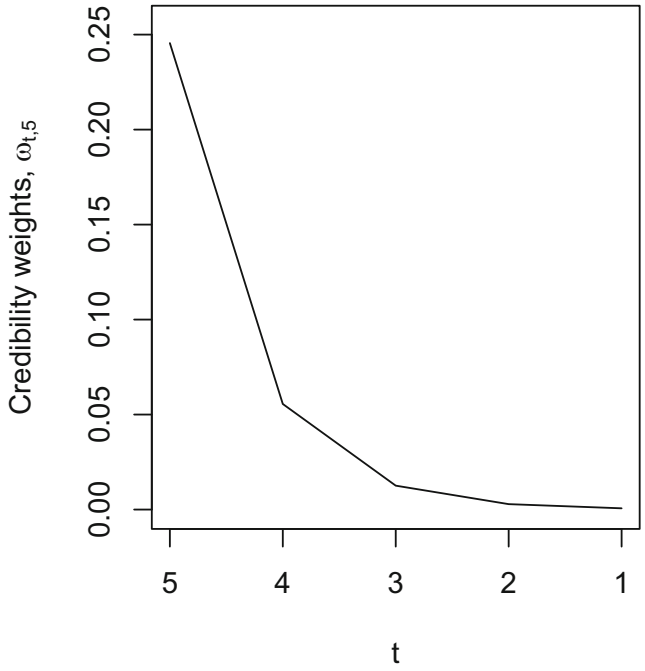


Figure 1: Credibility weights $(\omega_{t,5})_{t=1}^{5}$ showing the different seniority weightings of past claims ($\psi = 1, a_{1|0} = 3, \upsilon_t = 1, \Delta_t = \Delta = 0.3$ for $t = 1, \ldots, 5$).

the latent factors $(\Theta_{i,t})_{t\geq 1}$. As can be seen in equation (15), the state-space update from t to $t+1$ exhibits the following thinning property:

$$\mathrm{E}\left[\Theta_{i,t+1}^{-1}\middle|Y_{i,1:t}\right] = \Delta\,\mathrm{E}\left[\Theta_{i,t}^{-1}\middle|Y_{i,1:t}\right] + (1-\Delta).$$

The estimated value of $\Delta \simeq 0.3$ clearly distinguishes our model from the Bühlmann models (where $\Delta = 1$), highlighting the seniority weighting effect inherent in our approach. Specifically, our model ($\Delta \neq 1$) assigns different weights to observations based on their timing, unlike the Bühlmann models, where each observation contributes equally to the predictions.

Figure 1 illustrates the decay in the credibility weights defined in (26), starting from $\omega_{5,5}$ (corresponding to the most recent observation) to $\omega_{1,5}$ (corresponding to the oldest observation in this example). The figure demonstrates that credibility weights decrease rapidly as the age of claims increases. This is clearly different from the Bühlmann credibility, which equally weights all past claims.

Next, we analyzed the performance of these six models by an out-of-sample validation analysis. For this, we report the root-mean squared error (RMSE) and the average gamma deviance (GDEV) loss defined by

$$\mathrm{RMSE}(\widehat{\mu}_{1:M}, Y_{1:M}|\upsilon_{1:M}) = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(\widehat{\mu}_i\upsilon_i - Y_i)^2},$$

$$\mathrm{GDEV}(\widehat{\mu}_{1:M}, Y_{1:M}|\upsilon_{1:M}) = 2\sum_{i=1}^{M}\left(-\upsilon_i \log\left(\frac{Y_i}{\widehat{\mu}_i\upsilon_i}\right) + \frac{Y_i - \widehat{\mu}_i\upsilon_i}{\widehat{\mu}_i}\right),$$

where we set $-\upsilon_i \log(Y_i/(\widehat{\mu}_i\upsilon_i)) = 0$ if $Y_i = \upsilon_i = 0$, and we dropped the lower time index $T+1$ for convenience. The results are reported in Table 6.

On this dataset, the SSMs clearly have the best predictive performance, see Table 6. As stated above, the estimate $\Delta \simeq 0.3$ leads to a comparably fast decay (in seniority weighting) of past claims experience, making the SSMs more similar to the independent models ($\Delta = 0$) than to the Bühlmann models ($\Delta = 1$). This is generally confirmed by our results, as the independent models seem to perform better than the Bühlmann models on this dataset. However, we give a clear preference to the heterogeneous SSM.

1708945x, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjs.70024 by Ewha Womans University Library, Wiley Online Library on [21/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 6: Summary of out-of-sample validation on the real dataset.

	RMSE			GDEV		
	Independent	Bühlmann	SSM	Independent	Bühlmann	SSM
Homogeneous	11562.25	12688.83	10715.44	4512.37	4612.77	4185.06
Heterogeneous	10809.82	10940.15	10173.28	3324.28	3532.69	3118.87

From our real data example, we conclude that we give a clear preference to the observation-driven SSMs over the independent and the Bühlmann cases. In our example, we only studied eight development periods, that is, we are not in the long-term regime, which allows us to study the asymptotic variance behaviour of the inverse state-space process discussed in Section 4, but the different variance structures result in different seniority weightings in the predictive means of Corollary 2. The long-term variance behaviour starts to become interesting in financial and actuarial applications if one has high-frequency data; for example, the Generalized Smith–Miller model may form a suitable model class for modelling telematics car driving data; see Yanez et al. (2025) and Chan et al. (2024).

7 Conclusion

Observation-driven state-space models are widely used in time-series modelling and longitudinal data modelling due to their analytical tractability in many cases. For claim counts, Harvey and Fernandes (1989) introduced a Poisson-gamma observation-driven state-space model that is fully analytically tractable, and Ahn et al. (2025) later extended this model to accommodate flexible variance behaviours in the latent state-space dynamics.

For claim size modelling or positive continuous random response modelling, the gamma distribution is a natural and often preferred choice. Smith and Miller (1986) proposed the gamma-gamma observation-driven state-space model, offering a fully tractable modelling framework. However, a limitation of the Smith-Miller model is that its variance behaviour is constrained to be monotonically increasing. This article generalized the gamma-gamma observation-driven state-space model to allow for flexible variance behaviours, addressing this limitation.

Looking ahead, it would be valuable to develop a more rigorous classification of parameter-driven versus observation-driven state-space models. In particular, understanding the conditions under which a given response process $(Y_t)_{t\geq 1}$ admits both parameter-driven and observation-driven state-space representations would be of great interest. Naturally, this involves addressing the identifiability issues, which are not yet fully resolved.

The Poisson-gamma and gamma-gamma cases are part of the exponential dispersion family (EDF) with conjugate priors. A natural next step could also be to investigate whether our results can be extended to the entire class of EDF models with conjugate priors. Additionally, the gamma state-space process possesses several desirable properties, as exploited in Ahn et al. (2023b). It would be intriguing to determine the extent to which these properties are necessary to maintain analytical tractability.

Data Sharing

The data that support the findings of this study were derived from the Medical Expenditure Panel Survey (MEPS, `https://meps.ahrq.gov/mepsweb/index.jsp`), and the R codes are available at `https://github.com/ssauljin/Stationary_SM`.

Acknowledgements

The authors thank the Editor-in-Chief, Dr. Johanna Nešlehová, and two anonymous reviewers for their useful comments and suggestions that helped the authors to significantly improve this article.

Funding information

Jae Youn Ahn was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (2021R1A6A1A1003982, 2021R1A6A1A10039823, RS-2023-00217022, RS-2025-23524530) and an

1708945x, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjs.70024 by Ewha Womans University Library, Wiley Online Library on [21/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (RS-2022-00155966). Himchan Jeong was supported by the Natural Sciences and Engineering Research Council of Canada Alliance Catalyst Grant (R832535).

ORCID

Jae Youn Ahn: https://orcid.org/0000-0002-0492-2083
Himchan Jeong: https://orcid.org/0000-0001-5695-9964

A Simulation study

This appendix presents a simulation study to verify our modelling and estimation procedure. For the simulation study, we used the following data generation scheme. We assumed there were M distinct instances (insurance policies) that were observed for $T+1$ years. The records from the first T years were used for model fitting, and the records of year $T+1$ were used for the out-of-sample validation. The exposure at time $1 \le t \le T+1$ of instance $1 \le i \le M$, $\upsilon_{i,t}$, follows the following distribution:

$$\upsilon_{i,t} = N_{i,t} + B_{i,t} \ \in \ \mathbb{N}_0,$$

where $N_{i,t} \sim \mathcal{P}(0.2 \cdot (t+1))$, $B_{i,t} \sim \mathcal{B}(1.2 - 0.2t)$, and $N_{i,t}$ and $B_{i,t}$ are independent for all t and i. The claim severity at time t for instance i, $\mu_{i,t}$, is assumed to be known and selected by $\mu_{i,t} \sim \mathcal{U}(2000, 4000)$. This gives us the exogenous sequences $(\upsilon_{i,t}, \mu_{i,t})_{t=1}^{T+1}$ for all instances $1 \le i \le M$. Next, $(Y_{i,t}, \theta_{i,t})_{t=1}^{T+1}$ was generated under model 1, and assuming (20), which implies a variance stationary process. In the construction, we further assumed that $p_{i,t} = \kappa q_{i,t}$ for a constant $\kappa \ge 0$, which is equivalent to $\Delta_{i,t} \equiv \Delta$. Note that by construction, $Y_{i,t} = 0$ if and only if $\upsilon_{i,t} = 0$.

We generated 100 samples under the above data generation scheme with $M = 5000$ and $T = 5$, to estimate $a_{1|0}$, ψ, and Δ, and subsequently predict Y_6 from $Y_{1:5}$, $\upsilon_{1:6}$, and $\mu_{1:6}$. For comparison purposes, we study the same six models as in the real data example:

- Homogeneous independent: We assumed that $Y_{i,t} \big| Y_{i,1:t-1}$ follows a gamma distribution as in equation (2), with $\mu_{i,t} = \mu$ and $\theta_{i,t} = 1$ for $1 \le t \le 6$ and $1 \le i \le 5000$, where μ was estimated as the empirical weighted average of $(Y_{i,t})_{i,t}$ so that
$$\widehat{\mu} = \frac{\sum_{i=1}^{M}\sum_{t=1}^{T} Y_{i,t}}{\sum_{i=1}^{M}\sum_{t=1}^{T} \upsilon_{i,t}}.$$
- Homogeneous Bühlmann model: We assumed that $Y_{i,t} \big| Y_{i,1:t-1}$ follows a gamma distribution as in equation (2), $\mu_{i,t} = \mu$ and $\theta_{i,t} = \theta_i$ for $1 \le t \le 6$ and $1 \le i \le 5000$. Note that this is a special case of the next model (homogeneous SSM) with $\Delta = 1$.
- Homogeneous state-space model (SSM): We assumed that $Y_{i,t} \big| Y_{i,1:t-1}$ follows a gamma distribution as in equation (2), $\mu_{i,t} = \mu$ and we allow for an evolution of $\theta_{i,1:6}$ under model 1, assuming (20) holds.
- Heterogeneous independent model: We assumed that $Y_{i,t} \big| Y_{i,1:t-1}$ follows a gamma distribution as in equation (2), $\theta_{i,t} = 1$ for $1 \le t \le 6$ and $1 \le i \le 5000$ but we use given $\mu_{i,t}$'s.
- Heterogeneous Bühlmann model: We assumed that $Y_{i,t} \big| Y_{i,1:t-1}$ follows a gamma distribution as in equation (2), $\theta_{i,t} = \theta_i$ for $1 \le t \le 6$ and $1 \le i \le 5000$ but we use the given $\mu_{i,t}$'s. Note that this is a special case of the next model (heterogeneous SSM) with $\Delta = 1$.
- Heterogeneous SSM: We assumed that $Y_{i,t} \big| Y_{i,1:t-1}$ follows a gamma distribution as in equation (2) while we allow for an evolution of $\theta_{i,1:6}$ under model 1, assuming (20) holds, and we use the given $\mu_{i,t}$ for $1 \le t \le 6$ and $1 \le i \le 5000$.

Table A1 summarizes the estimation results of $a_{1|0}$, ψ, and Δ over the 100 simulations; both the mean estimates and the standard errors (in parentheses) are presented. Note that the homogeneous/heterogeneous independent models are again excluded from the comparison, as these models do not involve $a_{1|0}$, ψ and Δ (by assuming θ is constantly equal to one). It can be seen that the heterogeneous SSM accurately estimates the values of $a_{1|0}$, ψ, and Δ. In the case of the homogeneous SSM, these estimates are a little bit biased because it has a misspecified marginal mean structure.

1708945x, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjs.70024 by Ewha Womans University Library, Wiley Online Library on [21/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table A1: Summary of estimation procedure for true values $a_{1|0} = 3$ and $\psi = 1$.

	$\Delta = 0.5$			$\Delta = 1.0$		
	$\widehat{a}_{1\|0}$	$\widehat{\psi}$	$\widehat{\Delta}$	$\widehat{a}_{1\|0}$	$\widehat{\psi}$	$\widehat{\Delta}$
Homogeneous	5.5828	1.2222	1.0000	2.9460	1.0415	1.0000
Bühlmann	(0.2704)	(0.0114)	–	(0.1246)	(0.0088)	–
Homogeneous	2.7620	1.0151	0.4659	2.8364	1.0256	0.9750
SSM	(0.1273)	(0.0138)	(0.0238)	(0.1214)	(0.0104)	(0.0106)
Heterogeneous	5.8727	1.1858	1.0000	3.0151	1.0004	1.0000
Bühlmann	(0.2433)	(0.0108)	–	(0.1079)	(0.0083)	–
Heterogeneous	3.0279	1.0017	0.5027	2.9952	0.9977	0.9957
SSM	(0.1228)	(0.0135)	(0.0234)	(0.1077)	(0.0093)	(0.0065)

Table A2: Summary of out-of-sample validation with simulations ($a_{1|0} = 3, \psi = 1, \Delta = 0.5$).

	RMSE			GDEV		
	Independent	Bühlmann	SSM	Independent	Bühlmann	SSM
Homogeneous	6067.64	6133.54	5981.14	6680.86	6764.54	6499.17
	(724.26)	(690.46)	(701.65)	(180.72)	(189.24)	(175.27)
Heterogeneous	5974.26	6025.61	5878.33	6417.36	6473.82	6229.60
	(727.88)	(695.29)	(704.12)	(176.98)	(190.43)	(175.06)

Table A3: Summary of out-of-sample validation with simulations ($a_{1|0} = 3, \psi = 1, \Delta = 1.0$).

	RMSE			GDEV		
	Independent	Bühlmann	SSM	Independent	Bühlmann	SSM
Homogeneous	5962.36	5116.05	5118.61	6639.29	5206.91	5212.95
	(469.85)	(355.58)	(359.82)	(192.44)	(119.02)	(119.25)
Heterogeneous	5868.95	4939.25	4939.39	6371.98	4901.10	4901.98
	(471.67)	(351.47)	(351.82)	(186.97)	(116.44)	(116.38)

Tables A2 and A3 display the out-of-sample validation results. The means and standard errors (in parentheses) were obtained from the 100 iterations under the different values of Δ, which imply different state-space structures for the evolution of the latent factors $(\Theta_{i,t})_{t=1}^{T+1}$.

Table A2 considers the case $\Delta = 0.5$, and it shows that the true model (heterogeneous SSM) outperforms all other models in both measures. This is as expected, because the true model does not have any model error. In the case of $\Delta = 1.0$, Table A3, the heterogeneous Bühlmann model is the true model, so it shows the best predictive performance. Meanwhile, it is interesting to observe that the SSMs are capable of estimating the correct value of Δ and that their predictive performance is almost comparable to that of the true model (Bühlmann in this case). Therefore, we conclude that the proposed SSMs are flexible enough to find the true model in our simulation study.

B Proof of Lemma 1

This appendix gives the proof of Lemma 1. We start by proving the following three results:

$$\operatorname{var}[\Theta_{t+1}^{-1}] = \frac{a_{t+1|t}}{a_{t+1|t}-1}\operatorname{var}\left[\mathrm{E}\left[\Theta_{t+1}^{-1}\middle|Y_{1:t}\right]\right] + \frac{1}{a_{t+1|t}-1}, \tag{B1}$$

$$\operatorname{var}[\Theta_{t}^{-1}] = \frac{a_{t}}{a_{t}-1}\operatorname{var}\left[\mathrm{E}\left[\Theta_{t}^{-1}\middle|Y_{1:t}\right]\right] + \frac{1}{a_{t}-1}, \tag{B2}$$

1708945x, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjs.70024 by Ewha Womans University Library, Wiley Online Library on [21/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$\text{var}\left[\text{E}\left[\Theta_{t+1}^{-1}\middle|Y_{1:t}\right]\right] = \Delta_t^2 \text{var}\left[\text{E}\left[\Theta_t^{-1}\middle|Y_{1:t}\right]\right]. \tag{B3}$$

We use the tower property and mean stationarity, see Proposition 2, to show the first claim (B1) as follows:

$$\begin{aligned}
\text{var}[\Theta_{t+1}^{-1}] &= \text{E}\left[\text{var}[\Theta_{t+1}^{-1}|Y_{1:t}]\right] + \text{var}\left[\text{E}\left[\Theta_{t+1}^{-1}\middle|Y_{1:t}\right]\right] \\
&= \text{E}\left[\frac{b_{t+1|t}^2}{a_{t+1|t}^2(a_{t+1|t}-1)}\right] + \text{var}\left[\text{E}\left[\Theta_{t+1}^{-1}\middle|Y_{1:t}\right]\right] \\
&= \frac{1}{a_{t+1|t}-1}\text{E}\left[\text{E}\left[\Theta_{t+1}^{-1}\middle|Y_{1:t}\right]^2\right] + \text{var}\left[\text{E}\left[\Theta_{t+1}^{-1}\middle|Y_{1:t}\right]\right] \\
&= \frac{a_{t+1|t}}{a_{t+1|t}-1}\text{var}\left[\text{E}\left[\Theta_{t+1}^{-1}\middle|Y_{1:t}\right]\right] + \frac{1}{a_{t+1|t}-1}.
\end{aligned}$$

Similarly, we have for the second claim (B2)

$$\begin{aligned}
\text{var}[\Theta_t^{-1}] &= \text{E}\left[\text{var}[\Theta_t^{-1}|Y_{1:t}]\right] + \text{var}\left[\text{E}\left[\Theta_t^{-1}\middle|Y_{1:t}\right]\right] \\
&= \text{E}\left[\frac{b_t^2}{a_t^2(a_t-1)}\right] + \text{var}\left[\text{E}\left[\Theta_t^{-1}\middle|Y_{1:t}\right]\right] \\
&= \frac{1}{a_t-1}\text{E}\left[\text{E}\left[\Theta_t^{-1}\middle|Y_{1:t}\right]^2\right] + \text{var}\left[\text{E}\left[\Theta_t^{-1}\middle|Y_{1:t}\right]\right] \\
&= \frac{a_t}{a_t-1}\text{var}\left[\text{E}\left[\Theta_t^{-1}\middle|Y_{1:t}\right]\right] + \frac{1}{a_t-1}.
\end{aligned}$$

Finally, we have for the third claim (B3)

$$\text{var}\left[\text{E}\left[\Theta_{t+1}^{-1}\middle|Y_{1:t}\right]\right] = \text{var}[\Delta_t \text{E}\left[\Theta_t^{-1}\middle|Y_{1:t}\right] + (1-\Delta_t)] = \Delta_t^2 \text{var}\left[\text{E}\left[\Theta_t^{-1}\middle|Y_{1:t}\right]\right].$$

The lemma is now proved by merging (B1–B3) as follows:

$$\begin{aligned}
\text{var}[\Theta_{t+1}^{-1}] &= \frac{(p_t+q_t)a_t}{(p_t+q_t)a_t-1}\text{var}\left[\text{E}\left[\Theta_{t+1}^{-1}\middle|Y_{1:t}\right]\right] + \frac{1}{(p_t+q_t)a_t-1} \\
&= \Delta_t^2\frac{(p_t+q_t)a_t}{(p_t+q_t)a_t-1}\text{var}\left[\text{E}\left[\Theta_t^{-1}\middle|Y_{1:t}\right]\right] + \frac{1}{(p_t+q_t)a_t-1} \\
&= \Delta_t^2\frac{(p_t+q_t)a_t}{(p_t+q_t)a_t-1}\left(\frac{a_t-1}{a_t}\text{var}[\Theta_t^{-1}] - \frac{1}{a_t}\right) + \frac{1}{(p_t+q_t)a_t-1} \\
&= \Delta_t^2\frac{(p_t+q_t)(a_t-1)}{(p_t+q_t)a_t-1}\text{var}[\Theta_t^{-1}] + \frac{1-\Delta_t^2(p_t+q_t)}{(p_t+q_t)a_t-1}. \\
&= \frac{q_t^2}{p_t+q_t}\,\frac{a_t-1}{(p_t+q_t)a_t-1}\text{var}[\Theta_t^{-1}] + \left(1-\frac{q_t^2}{p_t+q_t}\right)\frac{1}{(p_t+q_t)a_t-1}.
\end{aligned}$$

This completes the proof. □

C Results on state-space updates in the Generalized Smith–Miller model

We come back to Remark 2. Since we are only interested in the marginal distribution of Y_{t+1}, given $Y_{1:t}$, the dependence structure of the state-space update in the observation-driven state-space model is not fully determined, see (8). One way of explicitly constructing the state-space update, which also allows us to simulate the Generalized Smith–Miller model, is to apply a thinning to the gamma state-space process from (3) to (4). This can be obtained from the following lemma.

1708945x, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjs.70024 by Ewha Womans University Library, Wiley Online Library on [21/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Lemma 6. *Consider three mutually independent random variables* Θ, B, *and* η *with*

$$\Theta \sim \gamma(1+a,b), \quad B \sim \text{Beta}(1+pa,(1-p)a), \quad \eta \sim \gamma(qa,pa+qb),$$

for $a>0$, $b>0$, $p\in[0,1]$, *and* $q\in\mathbb{R}_+$. *Then, we have the following results:*

i. $\Theta B \sim \gamma(1+pa,b)$.

ii. $\frac{\Theta B}{p} \sim \gamma(1+pa,pb)$.

iii. $\frac{\Theta B}{(pa+qb)/b}+\eta \sim \gamma(1+pa+qa,pa+qb)$.

iv. If $p+q\in(0,1)$, *consider* $B^*\sim \text{Beta}(1+pa+qa,(1-(p+q))a)$, *which is independent of* Θ, *such that*

$$\frac{\Theta B^*}{(pa+qb)/b} \sim \gamma(1+pa+qa,pa+qb).$$

For $p=1$, *we interpret* $B=1$, *a.s., which implies* $\Theta B=\Theta$, *a.s.*

Proof of Lemma 6. We compute the moment-generating function of $\log(\Theta)$. Choose $r>-(1+a)$, then

$$\mathrm{E}[\exp\{r\log(\Theta)\}]=\mathrm{E}\big[\Theta^r\big]=\frac{\gamma(1+a+r)}{\gamma(1+a)}b^{-r}.$$

For a beta-distributed random variable $B\sim\text{Beta}(c,d)$ we have for $r>-c$

$$\mathrm{E}[\exp\{r\log(B)\}]=\mathrm{E}[B^r]=\frac{\gamma(c+d)\gamma(c+r)}{\gamma(c+d+r)\gamma(c)}.$$

Set $c=1+pa$ and $d=(1-p)a$. This gives

$$\begin{aligned}
\mathrm{E}[\exp\{r\log(\Theta B)\}] &= \mathrm{E}\big[\Theta^r\big]\mathrm{E}[B^r] \\
&= \frac{\gamma(1+a+r)}{\gamma(1+a)}b^{-r}\cdot\frac{\gamma(c+d)\gamma(c+r)}{\gamma(c+d+r)\gamma(c)} \\
&= \frac{\gamma(1+a+r)}{\gamma(1+a)}b^{-r}\cdot\frac{\gamma(1+pa+(1-p)a)\gamma(1+pa+r)}{\gamma(1+pa+(1-p)a+r)\gamma(1+pa)} \\
&= \frac{\gamma(1+a+r)}{\gamma(1+a)}b^{-r}\cdot\frac{\gamma(1+a)\gamma(1+pa+r)}{\gamma(1+a+r)\gamma(1+pa)} \\
&= \frac{\gamma(1+pa+r)}{\gamma(1+pa)}b^{-r}.
\end{aligned}$$

This completes the proof of part (i). For part (ii), observe that

$$\mathrm{E}[\exp\{r\log(\Theta B/p)\}]=\mathrm{E}\big[\Theta^r\big]\mathrm{E}[(B/p)^r]=\frac{\gamma(1+pa+r)}{\gamma(1+pa)}(pb)^{-r}.$$

Part (iii) follows from the property that the sum of independent gamma random variables with the same rate parameter is also gamma-distributed with the summed shape parameters and the same rate parameter. Finally, we note that part (iv) can be proven similarly to part (ii). □

References

Ahn, J. Y., Jeong, H., and Lu, Y. (2021). On the ordering of credibility factors. *Insurance: Mathematics and Economics*, 101:626–638.

Ahn, J. Y., Jeong, H., and Lu, Y. (2023a). A simple Bayesian state-space approach to the collective risk models. *Scandinavian Actuarial Journal*, 2023(5):509–529.

1708945x, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjs.70024 by Ewha Womans University Library, Wiley Online Library on [21/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Ahn, J. Y., Jeong, H., Lu, Y., and Wüthrich, M. V. (2025). An observation-driven state-space count model for experience rating. *Insurance: Mathematics and Economics*, 125:103–149.

Anderson, B. D. and Moore, J. B. (2005). *Optimal Filtering*, Courier Corporation, North Chelmsford, Massachusetts.

Arulampalam, M. S., Maskell, S., Gordon, N., and Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188.

Bichsel, F. (1964). Erfahrungs-Tarifierung in der Motorfahrzeughaftpflicht-Versicherung. *Bulletin of the Swiss Association of Actuaries*, 1964:119–130.

Bolancé, C., Denuit, M., Guillén, M., and Lambert, P. (2007). Greatest accuracy credibility with dynamic heterogeneity: The Harvey-Fernandes model. *Belgian Actuarial Bulletin*, 7(1):14–18.

Bolancé, C., Guillén, M., and Pinquet, J. (2003). Time-varying credibility for frequency risk models: Estimation and tests for autoregressive specifications on the random effects. *Insurance: Mathematics and Economics*, 33(2):273–282.

Bühlmann, H. (1967). Experience rating and credibility. *ASTIN Bulletin - The Journal of the IAA*, 4(3):199–207.

Bühlmann, H. and Gisler, A. (2005). *A Course in Credibility Theory and Its Applications*, Springer Science & Business Media, Berlin and Heidelberg, Germany.

Bühlmann, H. and Straub, E. (1970). Glaubwürdigkeit für Schadensätze. *Bulletin of the Swiss Association of Actuaries*, 70(1):111–133.

Chan, I. W., Badescu, A. L., and Lin, X. S. (2024). Unsupervised detection of anomalous driving patterns using high resolution telematics time series data. arXiv:2412.08106.

Cox, D. R. (1981). Statistical analysis of time series: Some recent developments. *Scandinavian Journal of Statistics*, 8(2):93–115.

Doucet, A., Godsill, S., and Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10:197–208.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events: For Insurance and Finance*, Vol. 33, Springer Science & Business Media, Berlin and Heidelberg, Germany.

Gerber, H. U. and Jones, D. A. (1975a). Credibility formulas of the updating type. *Transactions of the Society of Actuaries*, 27(1):31–52.

Gerber, H. U. and Jones, D. A. (1975b). Credibility formulae with geometric weights. *Transactions of the Society of Actuaries*, 27:229–230.

Harvey, A. C. and Fernandes, C. (1989). Time series models for count or qualitative observations. *Journal of Business & Economic Statistics*, 7(4):407–417.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, Vol. 2, John Wiley & Sons, Hoboken, New Jersey.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974.

Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(4):619–656.

Lomax, K. S. (1954). Business failures: Another example of the analysis of failure data. *Journal of the American Statistical Association*, 49(268):847–852.

Pinquet, J. (2020a). Poisson models with dynamic random effects and nonnegative credibilities per period. *ASTIN Bulletin: The Journal of the IAA*, 50(2):585–618.

Pinquet, J. (2020b). Positivity properties of the ARFIMA (0, d, 0) specifications and credibility analysis of frequency risks. *Insurance: Mathematics and Economics*, 95:159–165.

Pinquet, J., Guillén, M., and Bolancé, C. (2001). Allowance for the age of claims in bonus-malus systems. *ASTIN Bulletin - The Journal of the IAA*, 31(2):337–348.

1708945x, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjs.70024 by Ewha Womans University Library, Wiley Online Library on [21/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Smith, R. and Miller, J. (1986). A non-Gaussian state space model and application to prediction of records. *Journal of the Royal Statistical Society: Series B*, 48(1):79–88.

Sundt, B. (1987). Credibility and old estimates. *Insurance: Mathematics and Economics*, 6(3):189–194.

Whitney, A. W. (1918). The theory of experience rating. *Proceedings of the Casualty Actuarial Society*, 4:274–292.

Yanez, J. S., Guillén, M., and Nielsen, J. P. (2025). Weekly dynamic motor insurance ratemaking with a telematics signals bonus-malus score. *ASTIN Bulletin - The Journal of the IAA*, 55(1):1–28.

Received 28 December 2024 — Accepted 1 July 2025

1708945x, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjs.70024 by Ewha Womans University Library, Wiley Online Library on [21/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License