EVOLUTIONARY HIERARCHICAL CREDIBILITY

BY

GREG TAYLOR

ABSTRACT

The hierarchical credibility model was introduced, and extended, in the 70s and early 80s. It deals with the estimation of parameters that characterize the nodes of a tree structure. That model is limited, however, by the fact that its parameters are assumed fixed over time. This causes the model's parameter estimates to track the parameters poorly when the latter are subject to variation over time. This paper seeks to remove this limitation by assuming the parameters in question to follow a process akin to a random walk over time, producing an evolutionary hierarchical model. The specific form of the model is compatible with the use of the Kalman filter for parameter estimation and forecasting. The application of the Kalman filter is conceptually straightforward, but the tree structure of the model parameters can be extensive, and some effort is required to retain organization of the updating algorithm. This is achieved by suitable manipulation of the graph associated with the tree. The graph matrix then appears in the matrix calculations inherent in the Kalman filter. A numerical example is included to illustrate the application of the filter to the model.

KEYWORDS

Credibility, credibility matrix, evolutionary model, graph matrix, hierarchy, Kalman filter, tree graph

1. INTRODUCTION

A hierarchical credibility model was introduced by Jewell (1975), and generalized by Taylor (1979). It is discussed further in Bühlmann and Jewell (1987). One of the applications mentioned in Taylor (1979) was workers compensation pricing, that, in certain jurisdictions, must be carried out for individual occupational categories that are arranged in a tree structure. Commercial insurance (fire, business interruption, etc.) is sometimes priced according to the same occupational structure. The tree structure was illustrated in Section 3 of Taylor (1979), and will be illustrated again in Section 2.1 of this paper.

Astin Bulletin 48(1), 339-374. doi: 10.1017/asb.2017.31

A similar example, possibly workers compensation but possibly some other class, would be concerned with the devolution of an organization's total premium to its cost centres, sub-centres, sub-sub-centres, etc. Commonly, the organization's insurer will quote just a total premium, and the organization will be left to conduct the devolution (see e.g., Comcare, 2017).

A further example might be provided by a consumer price index. Typically, such indexes are constructed by reference to a basket of goods from major categories (e.g., food, clothing, health, etc.), sub-categories, sub-sub-categories, etc. (see e.g., Australian Bureau of Statistics, 2011).

Bühlmann and Gisler (2006, Chapter 6) gave alternative applications to group accident insurance and industrial fire insurance, respectively.

The theoretical aspects of the subject were developed further by Sundt (1979, 1980), who placed a regression structure on observations at each level of the hierarchy, thereby generalizing the single-level regression credibility model of Hachemeister (1975). Norberg (1986) provided empirical Bayes estimators of the parameters of this model.

Alternative applications of hierarchical credibility models were suggested by Hesselager (1991), who applied them to loss reserving for a variety of claim types, and Ohlsson (2008), who applied them to motor pricing, though the hierarchy in each of these cases was relatively shallow.

Belhadj et al. (2009) presented three existing main sets of estimators of the variance components in the hierarchical credibility model literature using unified notations. They also studied their properties and compared their performance in numerical evaluation. These three sets of estimators include: iterative pseudo-estimators in Goovaerts and Hoogstad (1987) and Goovaerts et al. (1990), and two estimators that are somewhat more simply obtained by using known weights upon computing the estimator of a heterogeneity parameter in Ohlsson (2005) and Bühlmann and Gisler (2005).

Antonio et al. (2010) provided a multi-level analysis of intercompany claim counts. They developed hierarchical Poisson models that extend Jewell's hierarchical credibility model by incorporating risk factors in terms of explanatory variables. This aims to establish the connection between hierarchical credibility and multi-level statistics.

Piselis (2011) incorporated quantiles into the classical hierarchical credibility model of Jewell (1975) to provide the hierarchical credibility estimation of quantiles.

Ebrahimzadeh et al. (2013) developed a three-level credibility model for claims that incorporates common effects to allow for three sources of dependence: across portfolios, across individuals and across time within individuals. A general hierarchical credibility model is then derived for h levels of common effects.

All of these models are static. That is to say, their parameters are assumed fixed over the period of the data, or, to the extent that they involve random parameters, the distributions of those are assumed fixed.

Situations arise in which there is good cause to believe that parameters do not remain fixed over time. In the above case of workers compensation premium

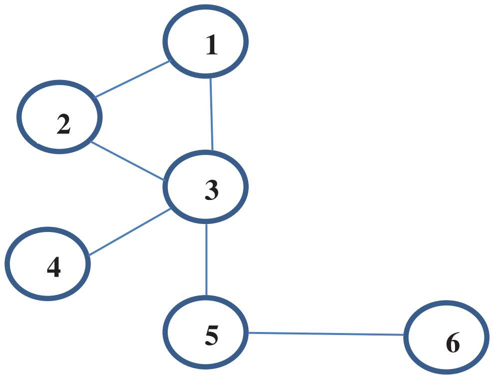

FIGURE 1: Diagrammatic representation of a graph. (Color online)

devolution, to take just one example, cost centres may respond to their devolved premiums by the successful implementation of mitigating risk controls, with resultant reduction in risk parameters.

In order to accommodate this sort of situation, it is necessary to formulate an evolutionary form of hierarchical model, in that the hierarchy itself remains unchanged over time, but the risk parameters associated with it are allowed to vary.

In this paper, this variation will take the form of a random walk, so that the evolutionary model is compatible with the Kalman filter. Full detail of this will appear in Sections 3 (model structure) and 4 (parameter estimates and forecasts).

Prior to this, Section 2 reviews the hierarchical framework itself and, later, Section 5 gives a numerical example of the credibility estimation at work.

2. HIERARCHICAL FRAMEWORK AND NOTATION

2.1. Graph theory

Subsequent sections will use some aspects of graph theory, and so a few fundamentals of the theory are reproduced below. These can be found, for example in Bondy and Murty (2008), Chen (1971) and Drmota (2009).

A graph $\mathcal{H}$ is an ordered pair $(V(\mathcal{H}), E(\mathcal{H}))$ consisting of a set $V(\mathcal{H})$ of vertices, or nodes, and a set $E(\mathcal{H})$, disjoint from $V(\mathcal{H})$, of edges, together with an incidence function $\Psi(\mathcal{H})$ that associates with each edge of $\mathcal{H}$ an unordered pair of (not necessarily distinct) nodes of $\mathcal{H}$. If e is an edge and u and v are nodes such that $\Psi(\mathcal{H})(e) = \{u, v\}$, then e is said to join u and v, and the nodes u and v are called the ends of e.

A graph is typically envisaged as a diagram such as Figure 1, in which the numbered circles represent the nodes and the lines joining them the edges.

https://doi.org/10.1017/asb.2017.31 Published online by Cambridge University Press

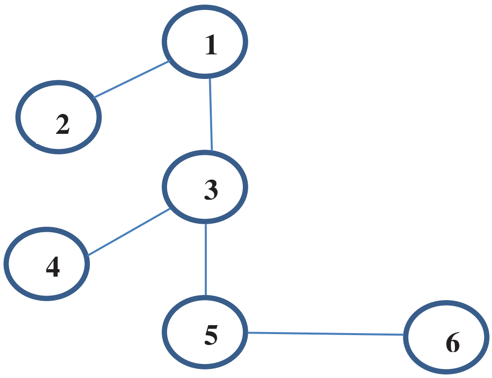

FIGURE 2: Diagrammatic representation of a tree. (Color online)

Let $v_1,\ v_2, \ldots, v_n$ be a finite sequence of nodes, and let $e_{12}, e_{23}, \ldots, e_{n-1,n}$ be a sequence of edges such that each $e_{i,i+1},\ i = 1, 2, \ldots, n-1$ joins v_i and v_{i+1}. Then $v_1,\ v_2, \ldots, v_n$ is called a path between v_1 and v_n. A path consisting of $n-1$ edges will be said to be of length $n-1$. For example, nodes 1,2,3,5,6 in Figure 1 form a path.

A tree is a graph in that any two nodes are connected by exactly one path. The graph in Figure 1 is not a tree because, for example, nodes 1 and 3 are connected by two paths, namely $\{1, 3\}$ and $\{1, 2, 3\}$. However, Figure 2 illustrates a tree.

A rooted tree is a tree in that one of the nodes is designated the root. In a rooted tree, a node that is connected to it on the path to the root is called the parent of that node. Every node except the root has a unique parent. A node of which v is the parent is a child of node v.

A child of a child of any node is a descendant of that node. By recursion, a child of any descendant of a node is itself a descendant of that node. Similarly, a parent of a parent of any node is an antecedent of that node. By recursion, a parent of any antecedent of a node is itself an antecedent of that node.

A childless node of the tree is called a leaf.

For example, if node 1 is chosen as a root in Figure 2, the tree there becomes a rooted tree. Node 3 becomes a parent of 4 and 5, and 4 and 5 are children of 3. Node 6 becomes a descendant of 3, and 3 an antecedent of 6. The leaves of the tree are 2, 4 and 6.

2.2. Hierarchical framework

Any tree may be considered as a hierarchy. Define the root of the tree to be level 0 of the hierarchy. Then, define level i of the hierarchy to consist of those nodes that are the children of nodes at level $i-1$, $i = 1, 2, \ldots$.

A tree, and its associated hierarchy, are called perfectly height balanced if all paths from root to leaf are of the same length (Choudum and Raman, 2009; Cha, 2012), i.e., all leaves are at the same level of the hierarchy, as illustrated in

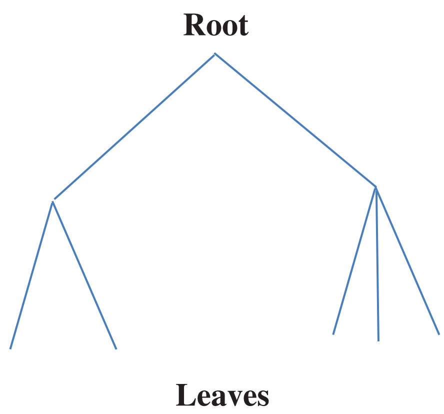

FIGURE 3: Illustration of a perfectly height balanced tree. (Color online)

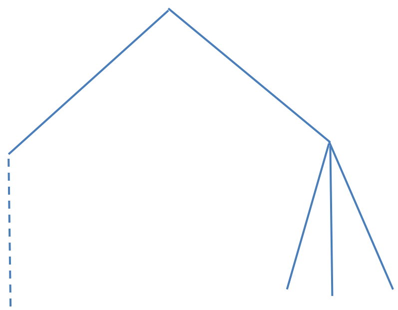

FIGURE 4: Extension of imperfectly to perfectly height-balanced tree. (Color online)

Figure 3. The hierarchy has levels 0,1,…,q if the number of edges separating root and leaf is q. This will be referred to as a q -hierarchy.

This paper will be concerned with only perfectly height balanced trees, but noting that any tree that is not of this type can be extended to perfect height balance, as illustrated in Figure 4, where the original tree consists of just the solid edges, and its extension includes the dashed one.

Informally, a q –hierarchy may be constructed as follows. A single node, the root, exists at level 0. For consistency with subsequent development, label it i_0.

Suppose the root has c children, and label them $1, 2, \ldots, c$. These constitute level 1 of the hierarchy. Denote the generic child at level 1 by i_1, and so the generic node at level 1 of the hierarchy by $i_0 i_1$.

Now, suppose that this node has $c(i_0 i_1)$ children. Level 2 of the hierarchy consists of such children, considered over all $i_1 = 1, 2, \ldots, c$. Denote the generic child of node $i_0 i_1$ at level 2 by i_2, and so the generic node at level 2 of the hierarchy by $i_0 i_1 i_2$. Continue the recursion to obtain a q -hierarchy.

https://doi.org/10.1017/asb.2017.31 Published online by Cambridge University Press

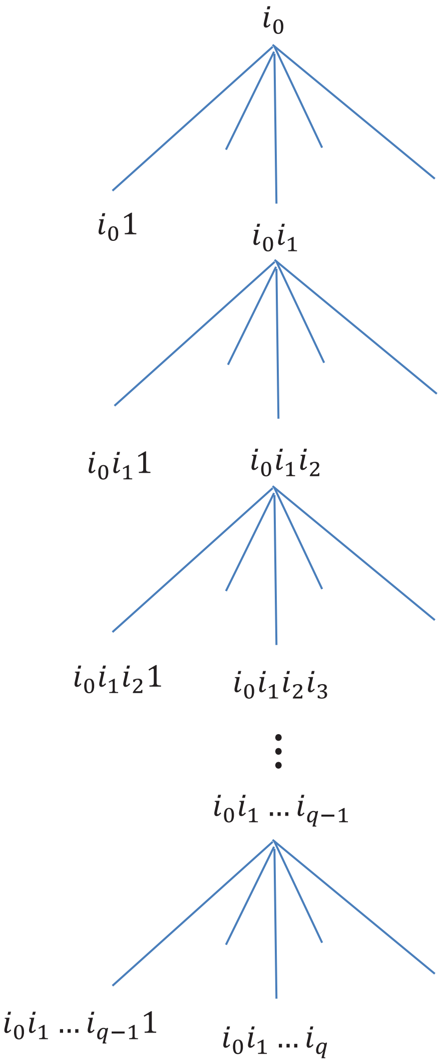

FIGURE 5: Illustration of a q –hierarchy. (Color online)

More formally, let the nodes at level $m(= 0, 1, \ldots, q)$ of a q –hierarchy be denoted $i_0 i_1 \ldots i_{m-1} i_m$, $i_m = 1, 2, etc.$, where, for fixed $i_0, i_1, \ldots, i_{m-1}$, these nodes (finite in number) are children of node $i_0 i_1 \ldots i_{m-1}$. Evidently, node i_0 is the root, and as there is only one of these, $i_0 = 1$.

Figure 5 illustrates a q –hierarchy with this nodal notation attached.

https://doi.org/10.1017/asb.2017.31 Published online by Cambridge University Press

Let the hierarchy illustrated in Figure 5 be denoted by $\mathcal{H}$. Let $\mathcal{H}_m$ denote the sub-hierarchy consisting of the nodes at levels $m-1$ and m, together with the edges between them.

2.3. Adjacency matrix of the hierarchy

The hierarchy $\mathcal{H}$ may be represented by its adjacency matrix $\Gamma(\mathcal{H})$. If all the nodes of $\mathcal{H}$ are placed in a specific order, $\Gamma(\mathcal{H})$ is an adjacency matrix whose (i, j) element is unity if the j -th node is a child of the i -th node, and is zero otherwise.

If the nodes are placed in dictionary order (i.e., 1,11,12,…,111,112,…,121,122, etc.), then the matrix is a block super-diagonal matrix, with q blocks on the super-diagonal, and with the m -th block the sub-matrix $\Gamma(\mathcal{H}_m)$, as follows:

$$\Gamma(\mathcal{H}) = \begin{bmatrix} 0 & \Gamma(\mathcal{H}_0) & \cdots & 0 \\ 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \Gamma\left(\mathcal{H}_{q-1}\right) \\ 0 & 0 & \cdots & 0 \end{bmatrix}. \tag{2.1}$$

The first row of $\Gamma(\mathcal{H}_m)$ contains the sub-matrix of the sub-hierarchy consisting of $i_0 i_1 \ldots i_m 1$ and its children, the second row sub-matrix of the sub-hierarchy consisting of $i_0 i_1 \ldots i_m 2$ and its children, etc.

It is evident, therefore, that the diagonal block $\Gamma(\mathcal{H}_m)$ is itself a block super-diagonal matrix of the form

$$\Gamma(\mathcal{H}_m) = \begin{bmatrix} 0 & u^T_{c(i_0 i_1 \ldots i_{m-1} 1)} & 0 & \cdots \\ 0 & 0 & u^T_{c(i_0 i_1 \ldots i_{m-1} 2)} & \\ \vdots & & 0 & \\ & & \ddots & u^T_{c(i_0 i_1 \ldots i_{m-1} c(i_0 i_1 \ldots i_{m-1}))} \\ & & & 0 \end{bmatrix}, \tag{2.2}$$

where u_n is an n -dimensional column vector with all components equal to unity, the upper T indicates matrix transposition, and $c(i_0 i_1 \ldots i_{m-1})$ is the number of children of node $i_0 i_1 \ldots i_{m-1}$.

It follows that the full matrix $\Gamma(\mathcal{H})$ is also a block super-diagonal matrix in which every block takes the form u_n^T for some n .

Elements of $\Gamma(\mathcal{H})$ or $\Gamma(\mathcal{H}_m)$ represent edges in their respective matrices, and will be labelled by the source and target nodes of those edges. Thus, for example, $\Gamma(\mathcal{H})_{i_0 i_1 \ldots i_m, j_0 j_1 \ldots j_n}$ will denote the element in the $i_0 i_1 \ldots i_m$ –th row and $j_0 j_1 \ldots j_n$ –th column, that records the incidence of an edge from the $i_0 i_1 \ldots i_m$ node to the $j_0 j_1 \ldots j_n$ node.

https://doi.org/10.1017/asb.2017.31 Published online by Cambridge University Press

It is evident from the properties of a tree that an edge exists between two given nodes if and only if target is a child of the source, i.e.,

$$\Gamma(\mathcal{H})_{i_0 i_1 \dots i_m, j_0 j_1 \dots j_n} = \delta_{m+1,n} \delta_{i_0 i_1 \dots i_m, j_0 j_1 \dots j_m}, \tag{2.3}$$

where δ_{pq} is the usual Kronecker delta, and $\delta_{i_0 i_1 \dots i_m, j_0 j_1 \dots j_m}$ is the multi-dimensional Kronecker delta:

$$\delta_{i_0 i_1 \dots i_m, j_0 j_1 \dots j_m} = \prod_{k=0}^{m} \delta_{i_k, j_k}. \tag{2.4}$$

It also follows from (2.3) that

$$\Gamma(\mathcal{H}_m)_{i_0 i_1 \dots i_m, j_0 j_1 \dots j_{m+1}} = \delta_{i_0 i_1 \dots i_m, j_0 j_1 \dots j_m}. \tag{2.5}$$

2.4. Multi-step matrix connections

The sub-matrix $\Gamma(\mathcal{H}_m)$ identifies all edges of the form $i_m i_{m+1}$. One may also construct an adjacency matrix $\Gamma(\mathcal{H}_{m:n})$, $m < n$ for all the edges $i_m i_{m+1} \dots i_n$ for fixed m, n . Note that $\Gamma(\mathcal{H}_m) \equiv \Gamma(\mathcal{H}_{m:m+1})$.

The following proposition is self-evident for a hierarchy.

Proposition 2.1. *For $m < n$, the $(i_0 i_1 \dots i_m, j_0 j_1 \dots j_n)$ -element of $\Gamma(\mathcal{H}_{m:n})$ is*

$$[\Gamma(\mathcal{H}_{m:n})]_{i_0 i_1 \dots i_m, j_0 j_1 \dots j_n} = \delta_{i_0 i_1 \dots i_m, j_0 j_1 \dots j_m}. \tag{2.6}$$

Lemma 2.2. *For any m, n, p with $m < p < n$,*

$$\Gamma(\mathcal{H}_{m:n}) = \Gamma\left(\mathcal{H}_{m:p}\right) \Gamma\left(\mathcal{H}_{p:n}\right). \tag{2.7}$$

The proof, along with others in this paper, is banished to Appendix A.

Corollary 2.3. *For any m, n with $m < n$,*

$$\Gamma(\mathcal{H}_{m:n}) = \Gamma(\mathcal{H}_m) \Gamma(\mathcal{H}_{m+1}) \dots \Gamma(\mathcal{H}_{n-1}).$$

Write (2.1) in the form

$$[\Gamma(\mathcal{H})]_{mn} = \delta_{m+1,n} \Gamma(\mathcal{H}_{m:m+1}), \quad m, n = 0, 1, \dots, q, \tag{2.8}$$

where, in this case, the subscripts label the blocks of $\Gamma(\mathcal{H})$.

The following lemma shows that the multi-step matrices $\Gamma(\mathcal{H}_{m:n})$ may be generated by taking powers of $\Gamma(\mathcal{H})$.

Lemma 2.4. *For $p = 1, 2, \dots, q$,*

$$[\Gamma^p(\mathcal{H})]_{mn} = \delta_{m+p,n} \Gamma\left(\mathcal{H}_{m:m+p}\right), \quad m, n = 0, 1, \dots, q. \tag{2.9}$$

https://doi.org/10.1017/asb.2017.31 Published online by Cambridge University Press

Also,

$$\Gamma^{q+1}(\mathcal{H}) = 0. \tag{2.10}$$

Thus, for $p = 1, 2, \ldots, q-1$,

$$\Gamma^{p}(\mathcal{H}) = \begin{bmatrix} 0 & \cdots & 0 & \Gamma\left(\mathcal{H}_{0:p}\right) & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \Gamma\left(\mathcal{H}_{1:1+p}\right) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & \Gamma\left(\mathcal{H}_{q-p:q}\right) \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}. \tag{2.11}$$

Lemma 2.5. *The following identity holds:*

$$[I - \Gamma(\mathcal{H})]^{-1} = \sum_{p=0}^{q} \Gamma^{p}(\mathcal{H}), \tag{2.12}$$

where the block structure of this matrix is

$$[I - \Gamma(\mathcal{H})]^{-1}_{mn} = \sum_{p=0}^{q} \delta_{m+p,n} \Gamma\left(\mathcal{H}_{m:m+p}\right), \tag{2.13}$$

with the convention that $\Gamma(\mathcal{H}_{m:m}) = I$.

3. Evolutionary hierarchical model

A q — hierarchical model will be created by placing a parameter vector β at each node of the q — hierarchy illustrated in Figure 5, and a set of observations at each leaf, as in Figure 6. The observations at level q are random variables y conditioned by the parameter vectors at that level, and each parameter vector at each level $q > 0$ is a random drawing from a distribution conditioned by the parameter vector at its parent node.

Let $\beta_{i_0 i_1 \ldots i_m}$ be the parameter vector associated with the $i_0 i_1 \ldots i_m$ node. Let $\beta_{(m)}$ denote the vector of all parameters at level m, obtained by stacking the vectors $\beta_{i_0 i_1 \ldots i_m}$, thus,

$$\beta_{(m)} = \begin{bmatrix} \beta_{i_0 i_1 \ldots i_{m-1} 1} \\ \beta_{i_0 i_1 \ldots i_{m-1} 2} \\ \vdots \end{bmatrix}.$$

https://doi.org/10.1017/asb.2017.31 Published online by Cambridge University Press

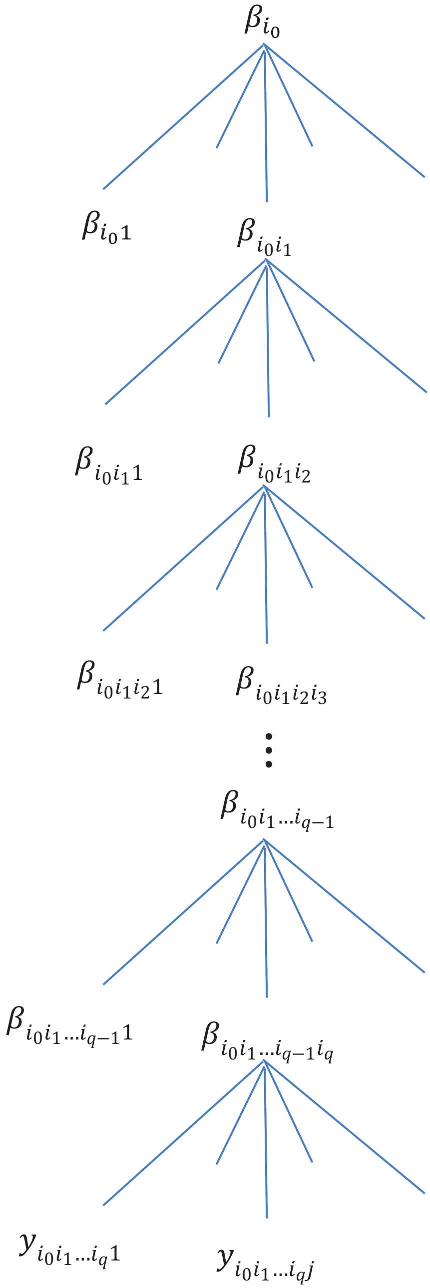

FIGURE 6: Illustration of a q -hierarchical model. (Color online)

https://doi.org/10.1017/asb.2017.31 Published online by Cambridge University Press

Further, let β denote the vector of all parameters:

$$\beta = \begin{bmatrix} \beta_{(0)} \\ \beta_{(1)} \\ \vdots \\ \beta_{(q)} \end{bmatrix}.$$

Further, $y_{i_0 i_1 \ldots i_q}$ denotes the vector of observations $y_{i_0 i_1 \ldots i_q j}$, and y denotes the vector obtained by stacking the $y_{i_0 i_1 \ldots i_q}$ in dictionary order with respect to node.

The above model is described in more formal terms as follows:

Model 3.1 (static). Consider a q -hierarchy $\mathcal{H}$, supplemented by parameters and observations that satisfy the following conditions:

a. A parameter vector $\beta_{i_0 i_1 \ldots i_m}$ is associated with node $i_0 i_1 \ldots i_m$ of the hierarchy.
b. The parameter β_{i_0} at the root of the tree is fixed.
c. For $m = 0, 1, \ldots, q-1$, the parameter vector $\beta_{i_0 i_1 \ldots i_m i_{m+1}}$ is a random drawing from some distribution determined by $\beta_{i_0 i_1 \ldots i_m}$.
d. At each of the hierarchy's terminal nodes $i_0 i_1 \ldots i_q$ there exists a sample of observations $y_{i_0 i_1 \ldots i_q j}$, $j = 1, 2, \ldots$ drawn from some distribution determined by $\beta_{i_0 i_1 \ldots i_q}$.
e. The random parameters and observations are subject to the following dependency structure:
   1. $\beta_{(m)} = W_{(m-1)} \beta_{(m-1)} + \zeta_{(m)}, m = 1, \ldots q$;
   2. $y = X\beta_{(q)} + \varepsilon$;

   where X is a design matrix, $W_{(m-1)}$ is some matrix compatible with the dimensions of $\beta_{(m-1)}$ and $\beta_{(m)}$, and $\zeta_{(m)}$, ε are random vectors, with ε independent of the $\zeta_{(m)}$, and
   3. $E[\zeta] = 0$, $E[\varepsilon] = 0$;
   4. $Var[\zeta] = \Lambda$, $Var[\varepsilon] = H$;

   where ζ is the vector obtained by stacking the $\zeta_{(m)}$.

The matrices $W_{(m)}$, $m = 0, 1, \ldots, q-1$ describe the way in which the parameter values are transmitted from one level of the hierarchy to the next, and will be referred to as transmission matrices.

This is a reasonably conventional hierarchical regression model. It is, in fact, essentially the same as the model of Sundt (1980), except that the latter places observations in a regression structure at each node of the hierarchy.

It is, however, a static model in the sense that, although the parameters are random, each is obtained by means of a single drawing from its distribution. The main purpose of the paper is to consider the situation in that observations are made at a sequence of epochs, with parameters evolving from one epoch to the next.

In recognition of the passage of time, all random quantities, and some non-random ones, are superscripted with a t, indicating time, e.g., $\beta^t_{i_0 i_1 \ldots i_m}$ is the value assumed by the parameter vector $\beta_{i_0 i_1 \ldots i_m}$ at time t . An evolutionary model is then created by retaining all features of Model 3.1, and adding further structure according to that parameters evolve over time. The model, written out in full, is as follows:

Model 3.2 (evolutionary). Consider a q -hierarchy $\mathcal{H}$, supplemented by parameters and observations that satisfy the following conditions. At each time $t = 0, 1, \ldots$:

a. A parameter vector $\beta^t_{i_0 i_1 \ldots i_m}$ is associated with node $i_0 i_1 \ldots i_m$ of the hierarchy.
b. For $m = 0, 1, \ldots, q-1$, the parameter vector $\beta^t_{i_0 i_1 \ldots i_m i_{m+1}}$ is a random drawing from some distribution determined by $\beta^t_{i_0 i_1 \ldots i_m}$.
c. At each of the hierarchy's terminal nodes $i_0 i_1 \ldots i_q$ there exists a sample of observations $y^t_{i_0 i_1 \ldots i_q j}$, $j = 1, 2, \ldots$ drawn from some distribution determined by $\beta^t_{i_0 i_1 \ldots i_q}$.
d. The observations are subject to the following dependency on parameters:
   1. $y^t = X^t \beta^t_{(q)} + \varepsilon^t$,
   where X^t is a design matrix, ε^t is a random vector, and
   2. $E[\varepsilon^t] = 0$, $Var[\varepsilon^t] = H^t$.

The parameter vector β^t evolves over time as follows: Define $\gamma^t_{(0)} = \beta^t_{(0)}$ and $\gamma^t_{(m)} = \beta^t_{(m)} - W_{(m-1)} \beta^t_{(m-1)}$, $m = 1, \ldots, q$. We assume that,

e. The parameter vector β^0 at $t = 0$ is random with known $E[\beta^0]$, $Var[\beta^0]$.
f. The parameters $\gamma^t_{(m)}$ evolve according to:
   3. $\gamma^t_{(m)} = \gamma^{t-1}_{(m)} + \zeta^t_{(m)}$, $m = 0, \ldots q$; $t = 1, 2, \ldots$,
   where $\zeta^t_{(m)}$ is a random vector, and
   4. $E[\zeta^t] = 0$, $Var[\zeta^t] = \Lambda^t$,
   and where ζ^t is the vector obtained by stacking the $\zeta^t_{(m)}$ and all ζ^t, ε^t, $t = 0, 1, 2, \ldots$ are mutually independent.

Remark. *The formulation in (f)(3) of the model is one in which each parameter evolves according to a random walk. There is no empirical evidence for this assumption. The long-term effect of the assumption will be for the parameters to drift apart indefinitely. A different model, such as mean-reverting could prevent that, though neither is there empirical evidence for this in all cases. This paper prefers the simpler random walk form, but with a caution over the model performance in prediction many steps ahead.*

An alternative form of assumption (3) is

$$\beta^t_{(m)} - \beta^{t-1}_{(m)} = W_{(m-1)} \left[\beta^t_{(m-1)} - \beta^{t-1}_{(m-1)} \right] + \zeta^t_{(m)}.$$

It is also possible to construct mappings between the vectors β^t and γ^t . By definition of $\gamma^t_{(m)}$,

$$\beta^t = \gamma^t + V\beta^t, \tag{3.1}$$

where V is the block matrix whose transpose is (transposition denoted by an upper T)

$$V^T = \begin{bmatrix} 0 & W^T_{(0)} & \cdots & 0 \\ 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & W^T_{(q-1)} \\ 0 & 0 & \cdots & 0 \end{bmatrix}. \tag{3.2}$$

This matrix can be recognized as having the same block form as the tree matrix $\Gamma(\mathcal{H})$ in (2.1). Lemmas 2.3 and 2.4 may be extended to the matrix V^T, the proofs running quite parallel to the proofs of those earlier lemmas.

Lemma 3.3. *For* $p = 1, 2, \ldots, q$,

$$[V^p]_{nm} = \delta_{m+p,n} W_{(m:m+p)}, \ m, n = 0, 1, \ldots, q,$$

where $W_{(m:m+p)} = W_{(m+p-1)} W_{(m+p-2)} \ldots W_{(m)}$.

Also,

$$V^{q+1} = 0.$$

Lemma 3.4. *The following identity holds:*

$$[I - V]^{-1} = \sum_{p=0}^{q} V^p,$$

where the block structure of this matrix is

$$[I - V]^{-1}_{nm} = \sum_{p=0}^{q} \delta_{m+p,n} W_{(m:m+p)},$$

with the convention that $W_{(m:m)} = I$.

By (3.1), the mappings between β^t and γ^t are

$$\gamma^t = (I - V)\,\beta^t, \beta^t = (I - V)^{-1}\gamma^t, \tag{3.3}$$

where V and $(I - V)^{-1}$ are given by (3.2) and Lemma 3.4, respectively.

https://doi.org/10.1017/asb.2017.31 Published online by Cambridge University Press

It will be helpful to re-formulate Model 3.2 entirely in terms of γ rather than β . To do so, note that, by (3.3) and Lemma 3.4,

$$\beta^t_{(q)} = \left[\, W_{(0:q)} \;\; W_{(1:q)} \;\; \cdots \;\; W_{(q:q)} \, \right] \gamma^t = W_{(*:q)}\gamma^t,$$

the matrix here being the last row of blocks in the block matrix $(I - V)^{-1}$

Therefore, re-write relation (d)(1) in the form

$$y^t = U^t \gamma^t + \varepsilon^t, \tag{3.4}$$

where

$$U^t = X^t W_{(*:q)} = \left[X^t \, W_{(0:q)} \;\; X^t W_{(1:q)} \;\; \cdots \;\; X^t W_{(q:q)} \right]. \tag{3.5}$$

4. Kalman filter forecast

4.1. Kalman filter

The Kalman filter was introduced by Kalman (1960). A description is found in Harvey (1989). Consider the following model.

Model 4.1. Suppose that:

a. The observation vector y^t is given by (3.4), called the observation equation or measurement equation.
b. Parameter evolution is given by relation (f)(3) of Model 3.2, called the system equation or transition equation.
c. The parameter vector γ^0 at $t = 0$ is random with known $E[\gamma^0]$, $Var[\gamma^0]$.
d. $\varepsilon^t \sim N(0, H^t)$, $\zeta^t \sim N(0, \Lambda^t)$, with all $\zeta^t, \varepsilon^t, t = 0, 1, 2, \ldots$ are mutually independent.

The model treated by Harvey is more general than this in two respects. First, the observation equation admits a deterministic drift vector. Second, the system equation admits both a deterministic drift vector and a linear transformation the parameter vector at each epoch. This additional generality is not required for present purposes.

The Kalman filter is a Minimum Mean Square Error (MMSE) Bayesian estimator. In the case of Model 4.1, involving normal errors in both observation and system equations, it also provides maximum a posteriori (MAP) estimators for the parameters of this model at each epoch, conditioned on past data, i.e., they maximize the likelihood of the Bayesian posterior distribution.

Let $\gamma^{t|s}$ and $P^{t|s}$ denote the MAP estimators of γ^t, $Var[\gamma^t - \gamma^{t|s}]$ given data $\{y^0, y^1, \ldots, y^s\}$. The filter comprises the following procedure for each $t = 1, 2, \ldots$:

1. Commence with estimate $\gamma^{t|t-1}$ and covariance matrix $P^{t|t-1}$ of the same dimension.
2. Calculate $F^t = U^t P^{t|t-1} (U^t)^T + H^t$.

https://doi.org/10.1017/asb.2017.31 Published online by Cambridge University Press

3. Calculate $K^t = P^{t|t-1}(U^t)^T(F^t)^{-1}$, called the Kalman gain matrix.
4. Update the matrix $P^{t|t-1}$ as follows:

$$P^{t+1|t} = P^{t|t-1} - P^{t|t-1}\left(U^t\right)^T\left(F^t\right)^{-1}U^t P^{t|t-1} + \Lambda^t.$$

5. Update the estimate $\gamma^{t|t-1}$ as follows: $\gamma^{t|t} = \gamma^{t|t-1} + K^t(y^t - U^t\gamma^{t|t-1})$.
6. Further update $\gamma^{t|t}$ as follows: $\gamma^{t+1|t} = \gamma^{t|t}$.

This procedure updates $\gamma^{t|t-1}$, $P^{t|t-1}$ to $\gamma^{t|t}$, $\gamma^{t+1|t}$, $P^{t+1|t}$. The procedure is initiated by setting $\gamma^{1|0} = E[\gamma^0]$, $P^{1|0} = Var[\gamma^0]$. Note that the estimators $\gamma^{t|t}$, $\gamma^{t+1|t}$ are linear in the data $\{y^0, y^1, \ldots, y^s\}$.

The update of the parameter vector estimate in step (5) takes the form of a linear combination of the existing estimate and the innovations vector, where the Kalman gain matrix serves as a weight applied to the latter. Steps (2) and (3) compute the gain matrix that is intimately related to the variation of the parmater vector over time, as is evident from step (4).

The entire algorithm specified by the Kalman filter, steps (1) to (6), is conveniently summarized in diagrammatic form by Figure 7.

The only difference between Models 3.2 and 4.1 is that the latter specifies normal distributions for observations and parameter variation over time, whereas the former does not. The Kalman filter is often applied in the absence of these distributional assumptions, but then its estimators are not generally MAP.

The advantage of the assumptions in the present case is that they ensure normal posterior distributions, whence MAP estimators are also MMSE unbiased linear Bayes, that is to say the same as credibility estimators. In other words, application of the Kalman filter will produce a credibility estimator at each epoch.

It is possible to express the Kalman updating formula from step (5) in forms more readily recognized as credibility forecasts:

$$\gamma^{t|t} = K^t y^t + \left(I - K^t U^t\right)\gamma^{t|t-1}, \tag{4.1}$$

$$y^{t|t} = U^t\gamma^{t|t} = Z^t y^t + \left(I - Z^t\right) y^{t|t-1}. \tag{4.2}$$

where $y^{t|t}$ denotes $E[y^t|y^0, y^1, \ldots, y^t] = E[U^t\gamma^t|y^0, y^1, \ldots, y^t]$ and the credibility matrix is

$$\begin{aligned} Z^t = U^t K^t &= U^t P^{t|t-1}\left(U^t\right)^T\left[U^t P^{t|t-1}\left(U^t\right)^T + H^t\right]^{-1} \\ &= U^t P^{t|t-1}\left(U^t\right)^T\left[H^t\right]^{-1}\left\{I + U^t P^{t|t-1}\left(U^t\right)^T\left[H^t\right]^{-1}\right\}^{-1}, \end{aligned} \tag{4.3}$$

on substitution of steps (2) and (3) of the filter in the penultimate step.

https://doi.org/10.1017/asb.2017.31 Published online by Cambridge University Press

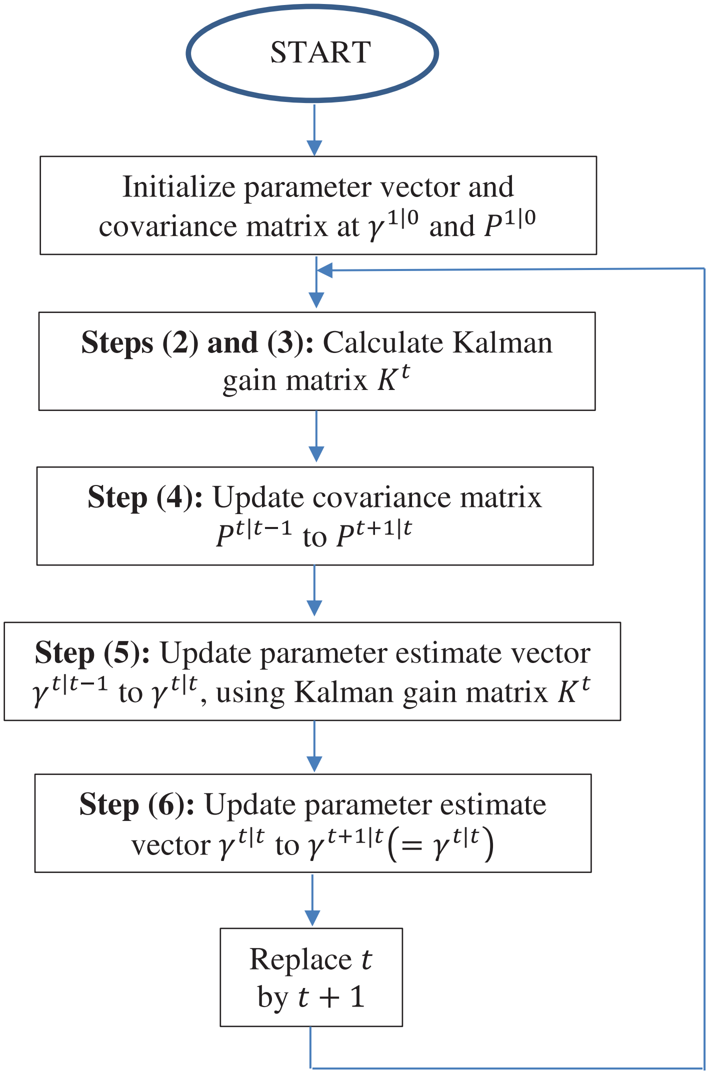

FIGURE 7: Diagrammatic representation of the Kalman filter. (Color online)

Similarly, the factor $K^t U^t$ in (4.1) may be re-expressed:

$$K^t U^t = P^{t|t-1}(U^t)^T\left[U^t P^{t|t-1}(U^t)^T + H^t\right]^{-1} U^t = P^{t|t-1}(U^t)^T\left[H^t\right]^{-1} \left\{I + U^t P^{t|t-1}(U^t)^T\left[H^t\right]^{-1}\right\}^{-1} U^t = P^{t|t-1}(U^t)^T\left[H^t\right]^{-1} U^t \left\{I + P^{t|t-1}(U^t)^T\left[H^t\right]^{-1} U^t\right\}^{-1}, \quad (4.4)$$

where the final step is justified by Proposition 4.2, immediately below.

Proposition 4.2. *If A, B be $m \times n$ and $n \times m$ matrices, respectively, then*

$$AB(I + AB)^{-1} = A(I + BA)^{-1} B = (I + AB)^{-1} AB,$$

provided that the inverse matrices in these expressions exist.

It is also possible to express $P^{t+1|t}$ briefly, using the definition of K^t and applying Proposition 4.2:

$$\begin{aligned} P^{t+1|t} &= P^{t|t-1} - K^t U^t P^{t|t-1} + \Lambda^t \\ &= P^{t|t-1} \left\{ I - \left(U^t\right)^T \left[H^t\right]^{-1} U^t \left\{ I + P^{t|t-1} \left(U^t\right)^T \left[H^t\right]^{-1} U^t \right\}^{-1} P^{t|t-1} \right\} + \Lambda^t \\ &= P^{t|t-1} \left[I - R^t \left(I + R^t\right)^{-1} \right] + \Lambda^t \\ &= P^{t|t-1} \left(I + R^t\right)^{-1} + \Lambda^t, \end{aligned} \tag{4.5}$$

where

$$R^t = \left(U^t\right)^T \left[H^t\right]^{-1} U^t P^{t|t-1}. \tag{4.6}$$

This form presents the updating of $P^{t|t-1}$ as a "scaling down" by a factor of $(I + R^t)$ for the passage of time, and then the addition of one period's additional parameter variance of Λ^t .

4.2. Decomposition of time-variation in parameter estimates

Consider the change in parameter estimate from $y^{t|t-1}$ to $y^{t|t}$ in (4.2):

$$y^{t|t} - y^{t|t-1} = Z^t \left(y^t - y^{t|t-1}\right), \tag{4.7}$$

with Z^t defined by (4.3), to which $P^{t|t-1}$ is seen to be a contributor.

Recall from Section 4.1 that $P^{t|t-1}$ is an estimate of $Var[\gamma^t - \gamma^{t|t-1} | y^0, y^1, \ldots, y^{t-1}]$, and that, by step (4) of the Kalman filter, it may be expressed in the form

$$P^{t|t-1} = Q^{t|t-1} + \Lambda^{t-1}, \tag{4.8}$$

where

$$Q^{t|t-1} = P^{t-1|t-2} - P^{t-1|t-2} \left(U^{t-1}\right)^T \left(F^{t-1}\right)^{-1} U^{t-1} P^{t-1|t-2}.$$

By condition (f)(4) of Model 3.2, Λ^{t-1} is the component of covariance introduced into (4.8) by the variation of parameters over time. If this were set to zero, then (4.8) would reduce to simply $P^{t|t-1} = Q^{t|t-1}$ and, in this sense, $Q^{t|t-1}$ may be viewed as that component of $P^{t|t-1}$ other than introduced by the variation of parameters over time, i.e., variation within the hierarchy.

It is of interest decompose (4.7) into these two components, as is done in Lemma 4.3.

Lemma 4.3. *Define*

$$Z_H^t = U^t Q^{t|t-1} \left(U^t\right)^T \left[H^t\right]^{-1} \left\{I + U^t Q^{t|t-1} \left(U^t\right)^T \left[H^t\right]^{-1}\right\}^{-1}, \tag{4.9}$$

$$Z_T^t = U^t \Lambda^{t-1} \left(U^t\right)^T \left[H^t\right]^{-1} \left\{I + U^t \Lambda^{t-1} \left(U^t\right)^T \left[H^t\right]^{-1}\right\}^{-1}. \tag{4.10}$$

Then, with Z^t defined by (4.3),

$$Z^t = Z_H^t \left\{I + Z_T^t \left[I - Z_T^t\right]^{-1} \left[I - Z_H^t\right]\right\}^{-1} + Z_T^t \left\{I + Z_H^t \left[I - Z_H^t\right]^{-1} \left[I - Z_T^t\right]\right\}^{-1}. \tag{4.11}$$

Remark 4.4. $Z_H^t = 0$ *when* $Q^{t|t-1} = 0$*, and* $Z_T^t = 0$ *when* $\Lambda^{t-1} = 0$ *. Thus,* $Z^t = Z_T^t$ *when* $Q^{t|t-1} = 0$*, and* $Z^t = Z_H^t$ *when* $\Lambda^{t-1} = 0$ *. In this sense,* Z_H^t *may be interpreted as the credibility matrix associated with parameter variation within the hierarchy at time* t*, and* Z_T^t *as the credibility matrix associated with parameter variation between times* $t-1$ *and* t *.*

Remark 4.5. *There are algebraic forms of the decomposition in Lemma 4.3 alternative to (4.11). However, that relation is preferred here since it takes the form* $Z^t = Z_H^t \times multiplier + Z_T^t \times multiplier$ *.*

4.3. Application to hierarchical forecast

4.3.1. *General case.* As already noted in Section 4.1, Model 4.1 to which the Kalman filter applies includes the hierarchical model 3.2 as a special case with U^t taking the form (3.5). one obtains a one-step-ahead forecast as

$$y^{t+1|t} = U^{t+1} \gamma^{t+1|t} = U^{t+1} \gamma^{t|t} = U^{t+1} \left[K^t y^t + \left(I - K^t U^t\right) \gamma^{t|t-1}\right], \tag{4.12}$$

where (3.4) has been used, then step (6) of the Kalman filter, followed by (4.1).

The estimates $\gamma^{t|t}$, $t = 1, 2, \ldots$ are obtained by repeated use of the Kalman loop set out in Section 4.1, after initiation as also set out there.

This is a straightforward procedure, consisting mainly of the matrix manipulation appearing in the loop. However, there is a commonly occurring special case, for that certain parts of the calculation simplify. These are discussed in the following sub-section.

4.3.2. *Parameter estimation.* As noted in condition (e) of Model 3.2, the initiating parameters $E[\beta^0]$, $Var[\beta^0]$ are assumed known. In addition, the model's covariance matrices Λ^t, relating to parameter variation over time, and H^t, relating to noise in observations, are assumed known.

Estimates of these will be required for implementation of the model. It is an unfortunate fact that, in practical situations, data for parameter estimation will often be unavailable. This will be particularly the case when:

https://doi.org/10.1017/asb.2017.31 Published online by Cambridge University Press

- the hierarchical filter is being implemented, and the relevant data have not previously been collected;
- the filter is in operation, and data have been collected, but over too short a duration for reliable parameter estimates to be made;
- the hierarchy has been subjected to structural changes over time, e.g., nodes shifted from one parent to another, preventing the collection of data within a framework that is time-consistent.

In these cases, informed guess work may be required.

However, in the event that the required data are available, estimation might be carried out in accordance with Durbin and Koopman (2012), where maximum likelihood estimation of the structural parameters Λ^t and H^t is discussed at some length (Chapter 7).

Durbin and Koopman (2012, Chapter 5) also consider initiation of the filter with the selection of values for $E[\beta^0]$, $Var[\beta^0]$. They cite Rosenberg (1973), De Jong (1991) and Koopman and Durbin (2003). The first of these uses maximum likelihood estimation, whereas the other two assume estimate $E[\beta^0]$ on the basis of indefinitely large $V[\beta^0]$.

4.3.3. *Special case.* Consider the case in which the design matrix $X^t = I$, the process covariance matrix H^t is diagonal, and

$$W_{(m)} = [\Gamma\,(\mathcal{H}_m)]^T, \tag{4.13}$$

i.e., by (2.2),

$$\left[W_{(m)}\right]_{i_0 i_1 \ldots i_{m+1},\, j_0 j_1 \ldots j_m} = \delta_{i_0 i_1 \ldots i_m,\, j_0 j_1 \ldots j_m}. \tag{4.14}$$

The choice of design matrix implies, by conditions (d)(1) and (2) of Model 3.2, that $E[y^t] = \beta^t_{(q)}$, i.e., the parameters to be estimated are simply means of the observations. As an example, $\beta^t_{(m)}$ might represent the vector of claim frequencies at level m, each of these claim frequencies is an evolving random perturbation of the frequency at its parent node, and the observations y^t are themselves claim frequencies, and therefore unbiased with respect to the mean claim frequency at level q .

The choice of design matrix also implies, by (3.5),

$$U^t = W_{(*:q)} = \left[\, W_{(0:q)}\;\; W_{(1:q)}\;\; \cdots\;\; W_{(q:q)} \,\right]. \tag{4.15}$$

Substitution of (4.14) in the definition of $\gamma^t_{(m)}$ in Model 3.2 yields

$$\left[\gamma^t_{(m)}\right]_{i_0 i_1 \ldots i_m} = \left[\beta^t_{(m)}\right]_{i_0 i_1 \ldots i_m} - \left[\beta^t_{(m)}\right]_{i_0 i_1 \ldots i_{m-1}},$$

which is, for fixed t, just the perturbation of a β parameter at a node from the β at its parent node. By condition (f)(3) of the same model, the perturbations are assumed to follow a stationary process.

https://doi.org/10.1017/asb.2017.31 Published online by Cambridge University Press

In this case, the matrix $(U^t)^T[H^t]^{-1}U^t$ that appears in (4.4) and (4.6) simplifies considerably. Denote this matrix by A^t . Since H^t is assumed diagonal, let $h^t_{i_0i_1\ldots i_m}$ denote the diagonal entry relating to the $i_0i_1\ldots i_m$ node.

By (4.15),

$$A^t = \left(W_{(*:q)}\right)^T\left[H^t\right]^{-1}W_{(*:q)}, \tag{4.16}$$

which is a block matrix with (m, n) block

$$A^t_{[mn]} = \left(W_{(m:q)}\right)^T\left[H^t\right]^{-1}W_{(n:q)} = \Gamma\left(\mathcal{H}_{(m:q)}\right)\left[H^t\right]^{-1}\left[\Gamma\left(\mathcal{H}_{(n:q)}\right)\right]^T, \tag{4.17}$$

by Corollary 2.3 and (4.13).

Lemma 4.6. *For $m \leq n$, the $(i_0i_1\ldots i_m, j_0j_1\ldots j_n)$ -element of the matrix $A^t_{[mn]}$ is*

$$A^t_{[mn]i_0i_1\ldots i_m, j_0j_1\ldots j_n} = \delta_{i_0i_1\ldots i_m, j_0j_1\ldots j_m}\sum_{k_{n+1}\ldots k_q} h^{-1}_{j_0j_1\ldots j_nk_{n+1}\ldots k_q}.$$

By (4.16), the matrix A^t is symmetric, and so matrix blocks $A^t_{[mn]}$, $m > n$ can be found as $A^t_{[mn]} = [A^t_{[nm]}]^T$. The lemma demonstrates that each block $A^t_{[mn]}$, and therefore the entire matrix A^t, may be generate just by taking defined sums of reciprocals of the process covariance matrix H^t .

These results can be useful in the application of the Kalman filter to the current special case. Substitution of (4.15) and (4.16) in (4.6), and then in (4.5), yields the following alternative form of that last relation:

$$P^{t+1|t} = P^{t|t-1}\left(I + A^t P^{t|t-1}\right)^{-1} + \Lambda^t, \tag{4.18}$$

with A^t obtained by means of Lemma 4.6.

5. Numerical example

A numerical example is now given in which, for manageability of presentation of the results, the hierarchy size is small, specifically $q = 2$, with only 10 terminal nodes. It is emphasized that, in practice, the dimension of the problem may be scaled up without difficulty.

The hierarchy considered is as follows:

Level 0: Single node 1.
Level 1: Nodes 11,12,13.
Level 2: Nodes 111,112,121,122,123,124,131,132,133,134.

The special case of Section 4.3.3, in which $X^t = I$, $W_{(m)} = [\Gamma(\mathcal{H}_m)]^T$, is illustrated. It is assumed that covariance matrices Λ^t, H^t are diagonal.

It is also assumed that the observations y^t consist of observed claim frequencies per exposure $E^t_{i_0i_1\ldots i_q}$, and that $Var[y^t_{i_0i_1\ldots i_q}] = E[y^t_{i_0i_1\ldots i_q}]/E^t_{i_0i_1\ldots i_q} =$

https://doi.org/10.1017/asb.2017.31 Published online by Cambridge University Press

TABLE 1

PARAMETERS FOR NUMERICAL EXAMPLE.

Node	Initial Values ($\boldsymbol{\beta}^0$) Mean	Initial Values ($\boldsymbol{\beta}^0$) Variance	Parameter Variance (Λ^t)
1	0.070	0.00005	0.00005
11	0.025	0.00003	0.00001
12	0.100	0.00030	0.00005
13	0.150	0.00070	0.00015
111	0.010	0.00002	0.00001
112	0.035	0.00015	0.00002
121	0.050	0.00040	0.00004
122	0.080	0.00090	0.00010
123	0.100	0.00150	0.00015
124	0.120	0.00250	0.00025
131	0.135	0.00300	0.00030
132	0.155	0.00400	0.00040
133	0.180	0.00500	0.00050
134	0.200	0.00650	0.00070

$\beta^t_{i_0 i_1 \ldots i_q} / E^t_{i_0 i_1 \ldots i_q}$. These are the diagonal elements of H^t, and they are estimated by $\beta^{t|t-1}_{i_0 i_1 \ldots i_q} / E^t_{i_0 i_1 \ldots i_q}$ for inclusion in step (2) of the Kalman filter.

This is a (rather small-scale) example of the case of pricing by occupational group, mentioned in Section 1. Here, level 0 represents the total risk pool under some form of insurance, say workers compensation. The nodes at level 1 represent major occupational groups (Agriculture, Manufacturing, Clerical, etc.), and the nodes at level 2 occupational sub-groups (under Manufacturing: Food Product Manufacturing, Wood Product Manufacturing, etc.).

The observed claim frequencies at the leaves are the empirical frequencies in respect of the occupational sub-groups. The objective is to apply the model to estimate the frequency parameters for the sub-groups.

Although the assumptions $X^t = I$, $W_{(m)} = [\Gamma(\mathcal{H}_m)]^T$ are restrictive, they apply quite naturally to the practical situation. Hence, even under these restrictive assumptions, the model solves a realistic problem of substance.

The parameters associated with each node are set out in Table 1, and fabricated claim data in Table 2.

In the latter case, it is assumed, for simplicity, that exposures do not change over time. With the exception of the three shaded rows, the entries in the table have been generated as random perturbations of the initial values. The three exceptional cases adjust this randomness as follows:

- Node 121: upward trend of 0.015 per period added;
- Node 124: flat reduction of 0.040 made in each period;
- Node 132: downward trend of 0.020 per period added;

https://doi.org/10.1017/asb.2017.31 Published online by Cambridge University Press

TABLE 2

CLAIM DATA FOR NUMERICAL EXAMPLE.

Node	Exposure ($E^t_{i_0 i_1 \ldots i_q}$)	Observed Claim Frequency at t = 1	2	3
111	40	0.007	0.013	0.007
112	35	0.030	0.038	0.043
121	300	0.062	0.094	0.097
122	100	0.081	0.088	0.079
123	500	0.120	0.064	0.136
124	100	0.093	0.053	0.081
131	301	0.150	0.143	0.132
132	50	0.172	0.136	0.093
133	25	0.111	0.188	0.094
134	20	0.248	0.171	0.195

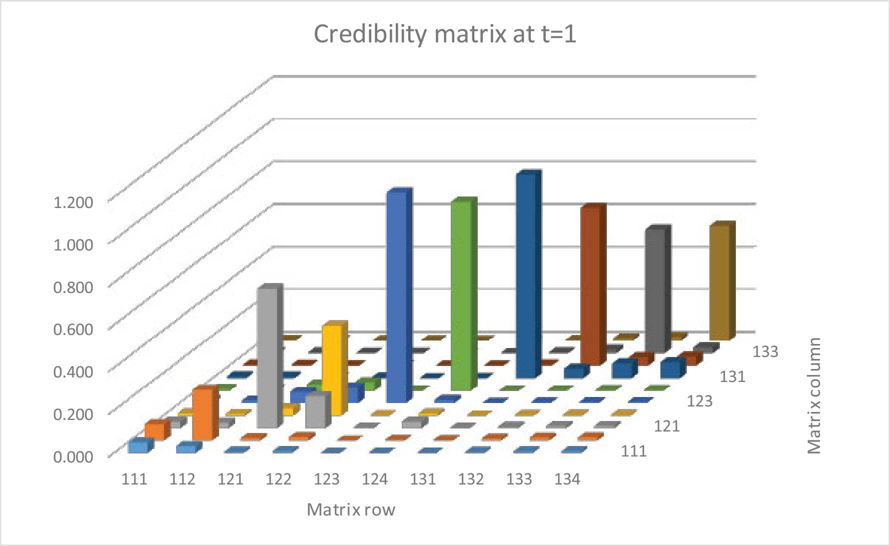

FIGURE 8: Credibility matrix Z^t at $t = 1$. (Color online)

Figure 8 displays diagrammatically the credibility matrix Z^t, defined by (4.3). Figures 9 and 10 display its decomposition into hierarchy and time components, given in (4.11). Appendix B contains the numerical detail.

It is seen that Z^t is dominated by its diagonal elements, i.e., the parameter estimate associated with any specific node is dominated by the observations at that node.

However, small contributions to credibility are made by off-diagonal elements. These are seen to be concentrated in the diagonal blocks that relate to

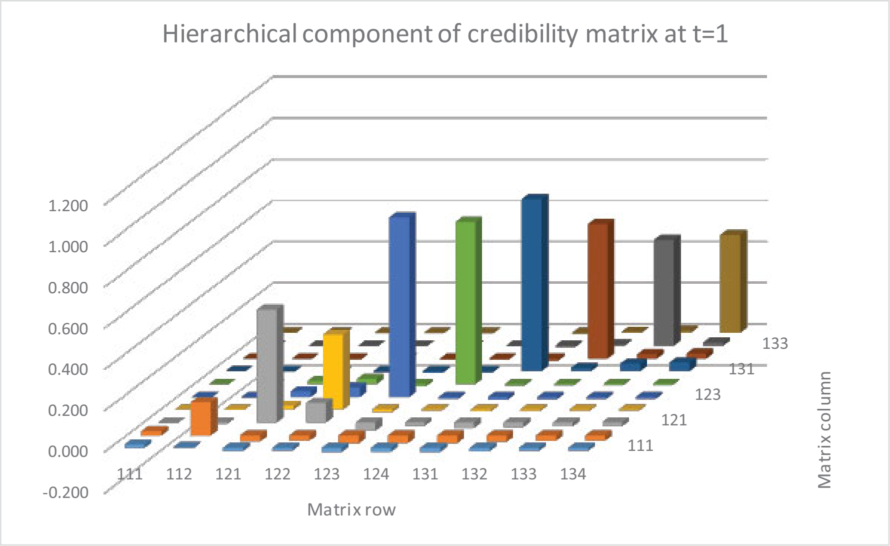

FIGURE 9: Hierarchy component of credibility matrix Z^t ($Z_H^t \times multiplier$) at $t = 1$. (Color online)

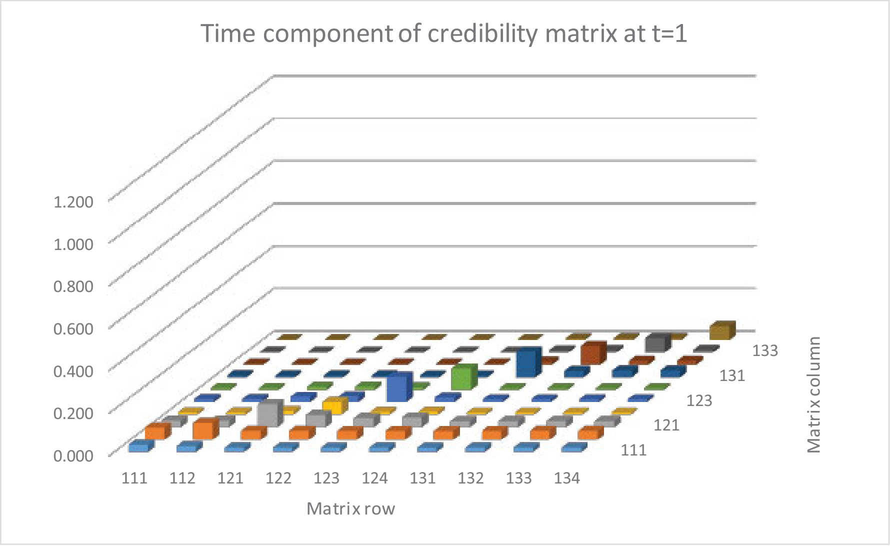

FIGURE 10: Time component of credibility matrix Z^t ($Z_T^t \times multiplier$) at $t = 1$. (Color online)

the three nodes 11,12,13 at level 1 of the hierarchy. That is, the parameter estimate associated with any specific node may be influenced in a minor way by observations at other nodes with the same parent.

https://doi.org/10.1017/asb.2017.31 Published online by Cambridge University Press

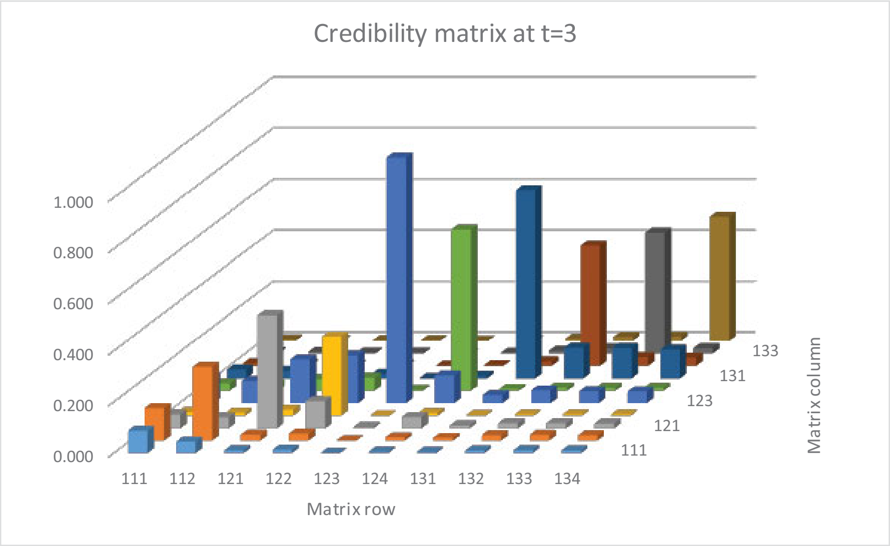

FIGURE 11: Credibility matrix Z^t at $t = 3$. (Color online)

It may also be noticed that the hierarchy component of credibility is dominant at $t = 1$, though this relation is reversed at later epochs (detail not given here).

Figure 11 illustrates the credibility matrix Z^t at $t = 3$, for comparison with the case $t = 1$ in Figure 8. Full numerical detail of all cases $t = 1, 2, 3$ is given in Appendix B.

Two features are evident from a comparison of this matrix with its counterpart at $t = 1$ (Figure 8). First, the larger diagonal elements of Z^t (nodes 12, 13) decrease between $t = 1$ and $t = 3$, whereas the smaller ones (node 11) increase. These effects occur because the sub-matrix of $P^{t+1|t}$ relating to nodes 12, 13 decreases with increasing t, whereas the sub-matrix relating to node 11 increases.

The second observable effect is that some of the off-diagonal elements of Z^t increase with t . This indicates that the extent to that the estimated frequency of a given terminal node is affected by its sibling nodes increases with the accumulation of information at those nodes.

Figure 12 displays the parameter estimates at all nodes, from the vectors $\beta^{t|t-1}$, as they evolve from the initial values at $t = 0$, through $t = 1, 2, 3$.

Observed and estimated claim frequencies are plotted for a selection of terminal nodes in Figures 13–17. Each of the figures plots the observed frequencies at $t = 1, 2, 3$, the prior frequency (estimate at $t = 0$), and the updated estimates at $t = 1, 2, 3$ according to (4.1). a confidence envelope is placed around the estimates.

Node	Estimated claim frequency and standard error (in parenthesis) for t=			
	0	1	2	3
1	0.070 (0.0071)	0.0703 (0.0096)	0.0644 (0.0109)	0.0743 (0.0118)
11	0.025 (0.0089)	0.0251 (0.0113)	0.0197 (0.0125)	0.0295 (0.0131)
12	0.100 (0.0187)	0.1034 (0.0166)	0.0951 (0.0171)	0.1185 (0.0175)
13	0.150 (0.0274)	0.1520 (0.0263)	0.1446 (0.0278)	0.1413 (0.0289)
111	0.010 (0.0100)	0.0100 (0.0125)	0.0048 (0.0138)	0.0144 (0.0146)
112	0.035 (0.0152)	0.0342 (0.0158)	0.0302 (0.0159)	0.0406 (0.0156)
121	0.050 (0.0274)	0.0580 (0.0189)	0.0650 (0.0174)	0.0912 (0.0167)
122	0.080 (0.0354)	0.0824 (0.0288)	0.0767 (0.0267)	0.0941 (0.0254)
123	0.100 (0.0430)	0.1198 (0.0164)	0.0680 (0.0164)	0.1331 (0.0161)
124	0.120 (0.0534)	0.0968 (0.0257)	0.0640 (0.0239)	0.0835 (0.0233)
131	0.135 (0.0612)	0.1491 (0.0257)	0.1429 (0.0255)	0.1363 (0.0257)
132	0.155 (0.0689)	0.1679 (0.0422)	0.1486 (0.0380)	0.1245 (0.0364)
133	0.180 (0.0758)	0.1428 (0.0549)	0.1545 (0.0502)	0.1268 (0.0445)
134	0.200 (0.0851)	0.2258 (0.0643)	0.1972 (0.0557)	0.1960 (0.0495)

FIGURE 12: Evolving parameter estimates.

https://doi.org/10.1017/asb.2017.31 Published online by Cambridge University Press

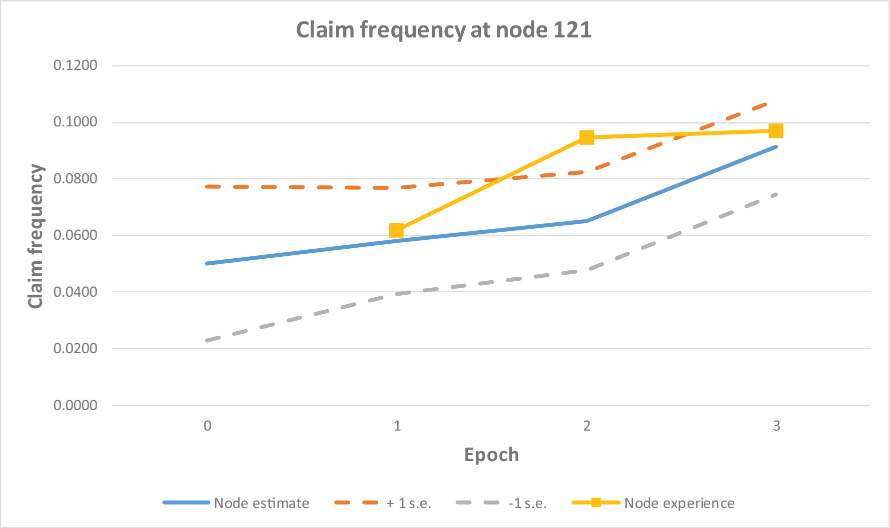

FIGURE 13: Estimation for node 121. (Color online)

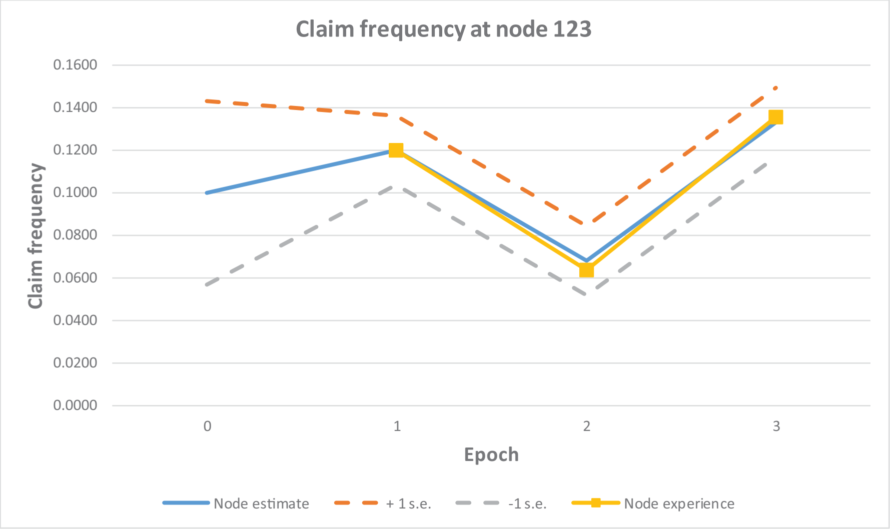

FIGURE 14: Estimation for node 123. (Color online)

Brief comments on the results are as follows:

- **Node 121.** High exposure, upward trend in claim frequency parameter. The estimates move upward over time.

https://doi.org/10.1017/asb.2017.31 Published online by Cambridge University Press

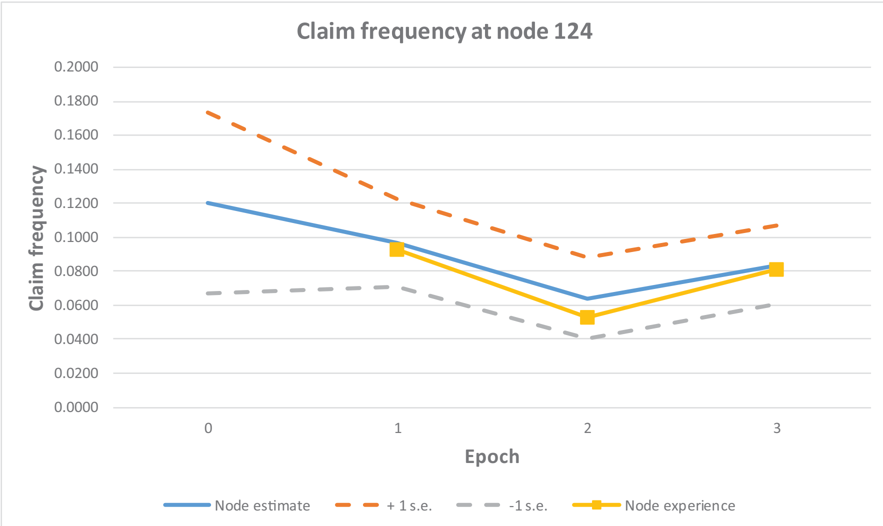

FIGURE 15: Estimation for node 124. (Color online)

- **Node 123.** High exposure, no trend in claim frequency parameter. Estimates follow experience closely.
- **Node 124.** Moderate exposure, no trend in claim frequency parameter, but prior over-estimated. Estimates lower than prior, in sympathy with experience.
- **Node 133.** Low exposure, downward trend in claim frequency parameter. Experience erratic, but on average lower than prior. Estimates display broadly declining trend.
- **Node 134.** Low exposure, no trend in claim frequency parameter. No trend in estimates.

6. CONCLUSION

An evolutionary hierarchical model has been formulated (Section 3), and estimates of it parameters constructed (Section 4). The parameter estimates yield forecasts of future observations.

The parameter estimates are obtained by application of the Kalman filter to the specific circumstance of the model. These estimates therefore update from one epoch to the next as further data are observed.

The application of the Kalman filter is conceptually straightforward, but the tree structure of the model parameters can be extensive, and some effort is required to retain organization of the updating algorithm. This is achieved by suitable manipulation of the adjacency matrix associated with the tree, as discussed in Section 2. That matrix can then be recruited to play its role in the matrix calculations inherent in the Kalman filter.

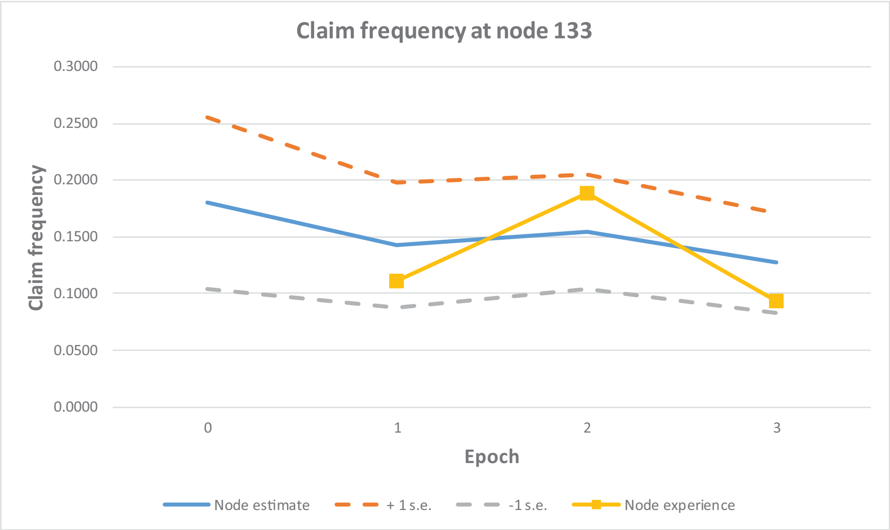

FIGURE 16: Estimation for node 133. (Color online)

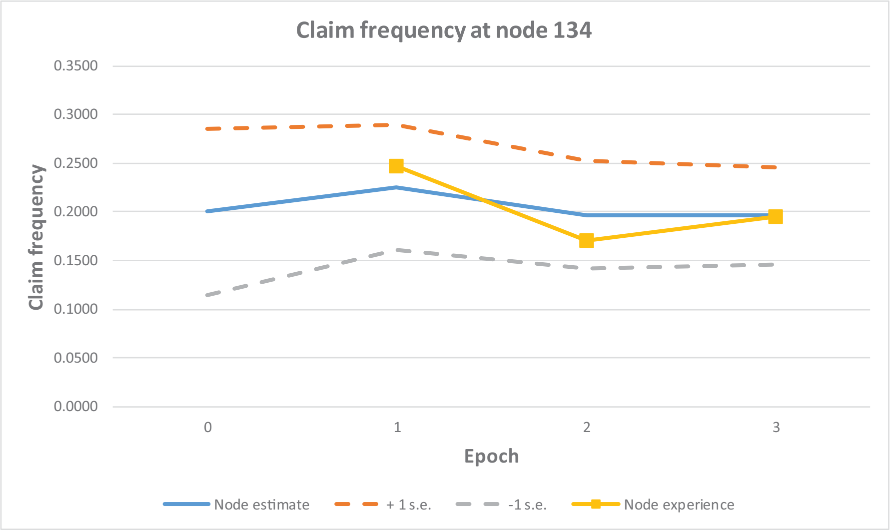

FIGURE 17: Estimation for node 134. (Color online)

It is also found that, in certain special cases, the book-keeping provided by these matrices can be highly simplified (Section 4.3.2).

The estimation and forecast algorithms provided in Section 4 consist essentially of a sequence of matrix calculations, and are simply implemented. In the case of the small-scale numerical example of Section 5, they were, in fact,

https://doi.org/10.1017/asb.2017.31 Published online by Cambridge University Press

implemented in Excel, though the exercise would have been less laborious if implemented by means of a genuine programming language such as R or C. Practicality would demand this in life-size problems.

The numerical example yields results that are intuitively explicable and reasonable. Section 1 mentions several contexts in that the evolutionary hierarchical model might be applicable. The results of the numerical example provide encouragement that the model would lead to reasonable parameter estimates and forecasts in those circumstances.

It should be noted that the model assumes normality of all distributions, both of observations and of random parameters. It would be possible, of course, and perhaps necessary in some contexts, to weaken this assumption. The use of conjugate pairs of distributions might enable the estimators to be extended although retaining their linear forms, but this speculation has been left for a future investigation.

Acknowledgements

The principal financial support for this research was provided by a grant from the Australian Actuaries Institute. The research was also supported under Australian Research Council's Linkage Projects funding scheme (project number LP130100723).

The views expressed herein are those of the author and are not necessarily those of either supporting body.

References

Antonio, K., Frees, E.W. and Valdez, E.A. (2010) A multilevel analysis of intercompany claim counts. *ASTIN Bulletin*, **40**(1), 151–177.

Australian Bureau of Statistics (2011) Consumer price index: Concepts, sources and methods. Catalogue number 6461.0. Canberra, Australia. Available at http://www.abs.gov.au/ausstats/abs@.nsf/mf/6461.0.

Belhadj, H., Goulet, V. and Ouellet, T. (2009) On parameter estimation in hierarchical credibility. *ASTIN Bulletin*, **39**(2), 495–514.

Bondy, J.A. and Murty, U.S.R. (2008) *Graph Theory with Applications*. Vol. 244 of Graduate Texts in Mathematics series. London: Springer-Verlag.

Bühlmann, H. and Gisler, A. (2006) *A Course in Credibility Theory and its Applications*. Berlin: Springer.

Bühlmann, H. and Jewell, W.S. (1987) Hierarchical credibility revisited. *Bulletin of the Swiss Association of Actuaries*, **87**, 35–54.

Cha, S.-H. (2012) On integer sequences derived from balanced k-ary trees. In: *Proceedings of American Conference on Applied Mathematics*, 377–381.

Chen, W.-K. (1971) *Applied Graph Theory*. Vol. 13 of North-Holland series in Applied Mathematics and Mechanics. Amsterdam: North-Holland Publishing Company.

Choudum, S. and Raman, I. (2009) Embedding height balanced trees and fibonacci trees in hypercubes. *Journal of Applied Mathematics and Computing*, **30**(1), 39–52.

Comcare (2017) Premium devolution. Available at https://www.comcare.gov.au/the_scheme/premium_paying_agencies/premium_devolution.

De Jong, P. (1991) The diffuse Kalman filter. *The Annals of Statistics*, **19**(2), 1073–1083.

DRMOTA, M. (2009) *Random Trees: An Interplay between Combinatorics and Probability*. Vienna: Springer Science & Business Media.
DURBIN, J. and KOOPMAN, S.J. (2012) *Time Series Analysis by State Space Methods*, 2nd Edition. Oxford, UK: Oxford University Press.
EBRAHIMZADEH, M., IBRAHIM, N.A., JEMAIN, A.A. and KILICMAN, A. (2013) Claim dependence induced by common effects in hierarchical credibility models. *Communications in Statistics-Theory and Methods*, **42**(18), 3373–3400.
GOOVAERTS, M. J., KAAS, R., VAN HEERWAARDEN, A.E. and BAUWELNCKX, T. (1990) *Effective Actuarial Methods*. Amsterdam: North-Holland Publishing Company.
GOOVAERTS, M.J. and HOOGSTAD, W.J. (1987) *Credibility Theory*. Vol. 4 of Surveys of actuarial studies. The Netherlands: Nationale-Nederlanden NV.
HACHEMEISTER, C.A. (1975) Credibility for regression models with application to trend. In *Credibility: Theory and Applications* (ed. P.M. Kahn), pp. 308–348. New York: Wiley.
HARVEY, A.C. (1989) *Forecasting, Structural Time Series and the Kalman Filter*. Cambridge, UK: Cambridge University Press.
HESSELAGER, O. (1991) Prediction of outstanding claims: A hierarchical credibility approach. *Scandinavian Actuarial Journal*, **1991**(1), 25–47.
JEWELL, W.S. (1975) The use of collateral data in credibility: A hierarchical model. *Research Memorandum 75-24 of the International Institute for Applied Systems Analysis*, Laxenburg, Austria: Institute for Applied Systems Analysis.
KALMAN, R.E. (1960) A new approach to linear filtering and prediction systems. *Journal of Basic Engineering*, **82**, 340–345.
KOOPMAN, S.J. and DURBIN, J. (2003) Filtering and smoothing of state vector for diffuse state-space models. *Journal of Time Series Analysis*, **24**(1), 85–98.
NORBERG, R. (1986) Hierarchical credibility: Analysis of a random effect linear model with nested classification. *Scandinavian Actuarial Journal*, **1986**(3–4), 204–222.
OHLSSON, E. (2005) Simplified estimation of structure parameters in hierarchical credibility. *36th ASTIN Colloquium*, Zurich.
OHLSSON, E. (2008) Combining generalized linear models and credibility models in practice. *Scandinavian Actuarial Journal*, **2008**(4), 301–314.
PITSELIS, G. (2011) On a hierarchical credibility model for quantiles. *40th ASTIN Colloquium*, Zurich, Available at http://www.actuaries.org/ASTIN/Colloquia/Madrid/Papers/Pitselis.pdf.
ROSENBERG, B. (1973) Random coefficients models: The analysis of a cross section of time series by stochastically convergent parameter regression. In: Annals of Economic and Social Measurement, *National Bureau of Economic Research*, **2**(4), 399–428.
SUNDT, B. (1979) A hierarchical credibility regression model. *Scandinavian Actuarial Journal*, **1979**(2–3), 107–114.
SUNDT, B. (1980) A multi-level hierarchical credibility regression model. *Scandinavian Actuarial Journal*, **1980**(1), 25–32.
TAYLOR, G.C. (1979) Credibility analysis of a general hierarchical model. *Scandinavian Actuarial Journal*, **1979**(1), 1–12.

GREG TAYLOR (Corresponding author)
UNSW Business School
Level 6, East Lobby
UNSW Business School Building E12
2052 Sydney, Australia.
Phone: +61 (0) 421 338 448.
E-mail: gregory.taylor@unsw.edu.au

https://doi.org/10.1017/asb.2017.31 Published online by Cambridge University Press

APPENDIX A

Proof of Lemma 2.2. Substitute Proposition 2.1 into the right side of (2.7):

$$\begin{aligned}\left[\Gamma\left(\mathcal{H}_{m:p}\right)\Gamma\left(\mathcal{H}_{p:n}\right)\right]_{i_0 i_1 \ldots i_m, j_0 j_1 \ldots j_n} &= \sum_{k_0 k_1 \ldots k_p}\left[\Gamma\left(\mathcal{H}_{m:p}\right)\right]_{i_0 i_1 \ldots i_m, k_0 k_1 \ldots k_p}\left[\Gamma\left(\mathcal{H}_{p:n}\right)\right]_{k_0 k_1 \ldots k_p, j_0 j_1 \ldots j_n} \\ &= \sum_{k_0 k_1 \ldots k_p}\delta_{i_0 i_1 \ldots i_m, k_0 k_1 \ldots k_m}\delta_{k_0 k_1 \ldots k_p, j_0 j_1 \ldots j_p} \\ &= \sum_{k_0 k_1 \ldots k_p}\delta_{i_0 i_1 \ldots i_m, k_0 k_1 \ldots k_m}\delta_{k_0 k_1 \ldots k_m, j_0 j_1 \ldots j_m}\delta_{k_m k_{\mathrm{m}+1} \ldots k_p, j_{\mathrm{m}} j_{\mathrm{m}+1} \ldots j_p},\end{aligned}$$

where (2.6) has been used in the second step.

Now for any given $j_{\mathrm{m}} j_{\mathrm{m}+1} \ldots j_p$, there is a single $k_m k_{\mathrm{m}+1} \ldots k_p$ for which $\delta_{k_m k_{\mathrm{m}+1} \ldots k_p, j_{\mathrm{m}} j_{\mathrm{m}+1} \ldots j_p} = 1$. For all other $k_m k_{\mathrm{m}+1} \ldots k_p$, $\delta_{k_m k_{\mathrm{m}+1} \ldots k_p, j_{\mathrm{m}} j_{\mathrm{m}+1} \ldots j_p} = 0$. Hence, the last relation reduces to

$$\begin{aligned}\left[\Gamma\left(\mathcal{H}_{m:p}\right)\Gamma\left(\mathcal{H}_{p:n}\right)\right]_{i_0 i_1 \ldots i_m, j_0 j_1 \ldots j_n} &= \sum_{k_0 k_1 \ldots k_m}\delta_{i_0 i_1 \ldots i_m, k_0 k_1 \ldots k_m}\delta_{k_0 k_1 \ldots k_m, j_0 j_1 \ldots j_m} \\ &= \delta_{i_0 i_1 \ldots i_m, j_0 j_1 \ldots j_m} = \Gamma(\mathcal{H}_{m:n})_{i_0 i_1 \ldots i_m, j_0 j_1 \ldots j_n} \text{ by (2.6).}\end{aligned}$$

■

Proof of Lemma 2.4. By (2.8), the lemma is true for $p = 1$. For larger values of p, proceed by induction. Assume that the lemma is true for p with $1 \le p < q$. Then

$$\left[\Gamma^{p+1}(\mathcal{H})\right]_{mn} = \sum_{k=0}^{q}[\Gamma(\mathcal{H})]_{mk}[\Gamma^p(\mathcal{H})]_{kn} = \sum_{k=0}^{q}\delta_{m+1,k}\Gamma(\mathcal{H}_{m:m+1})\,\delta_{k+p,n}\Gamma\left(\mathcal{H}_{k:k+p}\right)$$

[by the induction hypothesis (2.9)]

$$= \sum_{k=0}^{q}\delta_{m+1,k}\Gamma(\mathcal{H}_{m:m+1})\,\Gamma(\mathcal{H}_{k:n})$$

$$= \Gamma(\mathcal{H}_{m:k})\,\Gamma(\mathcal{H}_{k:n}) = \Gamma(\mathcal{H}_{m:n}),$$

by (2.7).

To prove (2.10), note that, for the case $p = q$, the term $\delta_{k+p,n}$ in the above development would require that $k = n - q$. Since $n \le q$, this can only produce non-negative k in the case $n = q$, $k = 0$. But then the term $\delta_{m+1,k}$ would require that $m = -1$ which cannot occur. Thus, the summation representing $[\Gamma^{q+1}(\mathcal{H})]_{mn}$ is vacuous, proving (2.10). ■

Proof of Lemma 2.5. The proof is a straightforward demonstration that

$$[I - \Gamma(\mathcal{H})][I - \Gamma(\mathcal{H})]^{-1} = I, \tag{A.1}$$

when (2.12) holds.

Note that, by (2.8),

$$[I - \Gamma(\mathcal{H})]_{mn} = \delta_{mn}I - \delta_{m+1,n}\Gamma(\mathcal{H}_{m:m+1}). \tag{A.2}$$

https://doi.org/10.1017/asb.2017.31 Published online by Cambridge University Press

By substitution of (2.12) and (A.2),

$$
\begin{aligned}
&\left[[I-\Gamma(\mathcal{H})][I-\Gamma(\mathcal{H})]^{-1}\right]_{mn} \\
&\quad = \sum_{k=0}^{q}\left\{[\delta_{mk}I-\delta_{m+1,k}\Gamma(\mathcal{H}_{m:m+1})]\sum_{p=0}^{q}\delta_{k+p,n}\Gamma\left(\mathcal{H}_{k:k+p}\right)\right\} \\
&\quad = \sum_{p=0}^{q}\delta_{m+p,n}\Gamma\left(\mathcal{H}_{m:m+p}\right)-\Gamma(\mathcal{H}_{m:m+1})\sum_{k=0}^{q}\sum_{p=0}^{q}\delta_{m+1,k}\delta_{k+p,n}\Gamma\left(\mathcal{H}_{k:k+p}\right) \qquad (A.3)\\
&\quad = \Gamma(\mathcal{H}_{m:n})-\Gamma(\mathcal{H}_{m:m+1})\sum_{k=0}^{n}\delta_{m+1,k}\Gamma(\mathcal{H}_{k:n}) \\
&\quad = \Gamma(\mathcal{H}_{m:n})-\Gamma(\mathcal{H}_{m:m+1})\Gamma(\mathcal{H}_{m+1:n})=0\ [\text{providedthat } m<n]
\end{aligned}
$$

$$= \Gamma(\mathcal{H}_{m:n})-\Gamma(\mathcal{H}_{m:n})=0, \tag{A.4}$$

by Proposition 2.1.

For the case $= n$, $\Gamma(\mathcal{H}_{m+1:n})=0$, and so (A.3) becomes

$$\left[[I-\Gamma(\mathcal{H})][I-\Gamma(\mathcal{H})]^{-1}\right]_{mm}=\Gamma(\mathcal{H}_{m:m})=I. \tag{A.5}$$

For the case $m > n$, all members of (A.3) are zero, and so it becomes

$$\left[[I-\Gamma(\mathcal{H})][I-\Gamma(\mathcal{H})]^{-1}\right]_{mm}=\Gamma(\mathcal{H}_{m:m})=I. \tag{A.6}$$

Finally, (A.4)–(A.6) amount to (A.1). ■

Proof of Lemma 4.3. To compute the required components of Z^t, first substitute (4.8) into (4.3):

$$
\begin{aligned}
Z^t &= U^t Q^{t|t-1}\left(U^t\right)^T\left[H^t\right]^{-1}\left\{I+U^t Q^{t|t-1}\left(U^t\right)^T\left[H^t\right]^{-1}+U^t\Lambda^{t-1}\left(U^t\right)^T\left[H^t\right]^{-1}\right\}^{-1} \\
&\quad + Q^{t|t-1}, \Lambda^{t-1},
\end{aligned} \tag{A.7}
$$

where a, b is defined as being equal to the previous member of the equation but with the roles of a, b reversed, i.e., a symmetrization operator.

For brevity, adopt the temporary notation:

$$A = U^t Q^{t|t-1}\left(U^t\right)^T\left[H^t\right]^{-1}, \tag{A.8}$$

$$B = U^t \Lambda^{t-1}\left(U^t\right)^T\left[H^t\right]^{-1}, \tag{A.9}$$

so that (A.7) may be expressed as

$$Z^t = A[I+A+B]^{-1}+A, B = A(I+A)^{-1}\left[I+B(I+A)^{-1}\right]^{-1}+A, B. \tag{A.10}$$

Now note that, by (4.9) and (4.10),

$$Z_H^t = A(I+A)^{-1}, \tag{A.11}$$

$$Z_T^t = B(I+B)^{-1}, \tag{A.12}$$

from which

$$I - Z_H^t = (I+A)^{-1}, \tag{A.13}$$

$$I - Z_T^t = (I+B)^{-1}, \tag{A.14}$$

https://doi.org/10.1017/asb.2017.31 Published online by Cambridge University Press

$$Z^t_H\left(I - Z^t_H\right)^{-1} = A, \tag{A.15}$$

$$Z^t_T\left(I - Z^t_T\right)^{-1} = B. \tag{A.16}$$

Substitution of (A.13)–(A.16) into (A.10) yields the lemma. ■

Proof of Lemma 4.6. By (4.17), and recognizing the assumed diagonal property of H^t,

$$\begin{aligned}
A^t_{[mn]i_0 i_1 \ldots i_m, j_0 j_1 \ldots j_n} &= \sum_{k_0 k_1 \ldots k_q} \left[\left[\Gamma\left(\mathcal{H}_{(m:q)}\right)\right]_{i_0 i_1 \ldots i_m, k_0 k_1 \ldots k_q} h^{-1}_{k_0 k_1 \ldots k_q} \left[\Gamma\left(\mathcal{H}_{(n:q)}\right)\right]_{j_0 j_1 \ldots j_n, k_0 k_1 \ldots k_q} \right] \\
&= \sum_{k_0 k_1 \ldots k_q} \left[\delta_{i_0 i_1 \ldots i_m, k_0 k_1 \ldots k_m} h^{-1}_{k_0 k_1 \ldots k_q} \delta_{j_0 j_1 \ldots j_n, k_0 k_1 \ldots k_n} \right] \\
&= \delta_{i_0 i_1 \ldots i_m, j_0 j_1 \ldots j_m} \sum_{k_{n+1} \ldots k_q} \left[h^{-1}_{i_0 i_1 \ldots i_m j_{m+1} \ldots j_n k_{n+1} \ldots k_q} \right],
\end{aligned}$$

where the second step has made use of . The lemma then follows. ■

APPENDIX B

The following figures evaluate the credibility matrix Z^t, defined by (4.3), and its decomposition given in (4.9)–(4.11). Partitions are inserted in each table to identify the diagonal block matrices that relate to the three nodes 11,12,13 at level 1 of the hierarchy. A row total for a particular node indicates the total credibility of all observations on the estimated frequency for that node. Figures B6 and B7 give the credibility matrix Z^t for $t = 2$ and $t = 3$ corresponding to $t = 1$ in Figure B1.

Node	Matrix entry for node										**Row total**
	111	112	121	122	123	124	131	132	133	134	
111	0.049	0.076	0.030	0.009	0.014	0.008	0.008	0.005	0.003	0.002	0.204
112	0.031	0.237	0.025	0.008	0.012	0.006	0.007	0.004	0.002	0.002	0.335
121	0.006	0.012	0.654	0.033	0.053	0.028	0.002	0.001	0.001	0.001	0.790
122	0.008	0.016	0.150	0.421	0.072	0.039	0.003	0.002	0.001	0.001	0.712
123	0.000	0.000	0.003	0.001	0.988	0.001	0.000	0.000	0.000	0.000	0.994
124	0.002	0.003	0.029	0.009	0.014	0.885	0.001	0.000	0.000	0.000	0.943
131	0.001	0.002	0.001	0.000	0.001	0.000	0.956	0.005	0.003	0.002	0.970
132	0.004	0.009	0.006	0.002	0.003	0.002	0.044	0.739	0.016	0.011	0.837
133	0.007	0.014	0.009	0.003	0.005	0.002	0.069	0.039	0.580	0.018	0.746
134	0.007	0.015	0.010	0.003	0.005	0.003	0.075	0.042	0.026	0.539	0.725

FIGURE B1: Total credibility matrix Z^t at $t = 1$. (Color online)

https://doi.org/10.1017/asb.2017.31 Published online by Cambridge University Press

Node	Matrix entry for node										Row total
	111	112	121	122	123	124	131	132	133	134	
111	0.017	0.023	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.039
112	0.009	0.173	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.183
121	0.000	0.000	0.618	0.030	0.050	0.026	0.000	0.000	0.000	0.000	0.725
122	0.000	0.000	0.135	0.388	0.068	0.036	0.000	0.000	0.000	0.000	0.627
123	0.000	0.000	0.003	0.001	0.987	0.001	0.000	0.000	0.000	0.000	0.992
124	0.000	0.000	0.028	0.008	0.014	0.873	0.000	0.000	0.000	0.000	0.923
131	0.000	0.000	0.000	0.000	0.000	0.000	0.950	0.004	0.003	0.002	0.959
132	0.000	0.000	0.000	0.000	0.000	0.000	0.042	0.715	0.014	0.010	0.782
133	0.000	0.000	0.000	0.000	0.000	0.000	0.064	0.035	0.551	0.016	0.666
134	0.000	0.000	0.000	0.000	0.000	0.000	0.069	0.038	0.023	0.508	0.639

FIGURE B2: Credibility matrix Z_H^t at $t = 1$. (Color online)

Node	Matrix entry for node										Row total
	111	112	121	122	123	124	131	132	133	134	
111	0.029	0.055	0.055	0.013	0.120	0.034	0.046	0.011	0.005	0.003	0.371
112	0.023	0.081	0.053	0.012	0.117	0.033	0.045	0.011	0.005	0.003	0.384
121	0.010	0.025	0.217	0.026	0.241	0.068	0.028	0.006	0.003	0.002	0.627
122	0.011	0.026	0.115	0.089	0.253	0.072	0.029	0.007	0.003	0.002	0.608
123	0.001	0.003	0.014	0.003	0.914	0.009	0.004	0.001	0.000	0.000	0.951
124	0.007	0.016	0.072	0.017	0.158	0.461	0.018	0.004	0.002	0.001	0.756
131	0.004	0.011	0.014	0.003	0.030	0.009	0.734	0.021	0.009	0.006	0.841
132	0.010	0.024	0.031	0.007	0.068	0.019	0.203	0.247	0.021	0.014	0.645
133	0.011	0.026	0.034	0.008	0.075	0.021	0.226	0.053	0.135	0.016	0.605
134	0.011	0.027	0.035	0.008	0.076	0.021	0.228	0.053	0.024	0.120	0.602

FIGURE B3: Credibility matrix Z_T^t at $t = 1$. (Color online)

Node	Matrix entry for node										Row total
	111	112	121	122	123	124	131	132	133	134	
111	0.015	0.020	-0.001	0.000	-0.001	0.000	0.000	0.000	0.000	0.000	0.032
112	0.004	0.159	-0.006	-0.002	-0.003	-0.001	-0.002	-0.001	-0.001	0.000	0.149
121	-0.015	-0.030	0.547	0.017	0.027	0.014	-0.006	-0.003	-0.002	-0.001	0.548
122	-0.013	-0.026	0.094	0.362	0.045	0.024	-0.005	-0.003	-0.002	-0.001	0.475
123	-0.020	-0.040	-0.039	-0.012	0.869	-0.010	-0.008	-0.004	-0.003	-0.002	0.731
124	-0.018	-0.038	-0.016	-0.005	-0.008	0.785	-0.007	-0.004	-0.003	-0.002	0.685
131	-0.019	-0.039	-0.026	-0.008	-0.013	-0.007	0.835	-0.010	-0.006	-0.004	0.704
132	-0.016	-0.032	-0.022	-0.007	-0.010	-0.006	0.015	0.651	0.005	0.004	0.584
133	-0.013	-0.028	-0.019	-0.006	-0.009	-0.005	0.037	0.021	0.513	0.009	0.502
134	-0.013	-0.026	-0.018	-0.005	-0.009	-0.005	0.043	0.024	0.015	0.475	0.481

FIGURE B4: Hierarchy component of credibility matrix Z^t ($Z_H^t \times multiplier$) at $t = 1$. (Color online)

Node	Matrix entry for node										Row total
	111	112	121	122	123	124	131	132	133	134	
111	0.033	0.056	0.031	0.009	0.015	0.008	0.009	0.005	0.003	0.002	0.172
112	0.027	0.078	0.030	0.009	0.015	0.008	0.009	0.005	0.003	0.002	0.186
121	0.020	0.042	0.108	0.016	0.026	0.014	0.008	0.005	0.003	0.002	0.243
122	0.021	0.042	0.056	0.059	0.027	0.014	0.008	0.005	0.003	0.002	0.237
123	0.020	0.041	0.042	0.013	0.119	0.011	0.008	0.004	0.003	0.002	0.263
124	0.020	0.041	0.045	0.014	0.022	0.099	0.008	0.004	0.003	0.002	0.258
131	0.020	0.040	0.027	0.008	0.013	0.007	0.120	0.015	0.009	0.007	0.266
132	0.020	0.041	0.028	0.008	0.013	0.007	0.029	0.088	0.010	0.007	0.253
133	0.020	0.042	0.028	0.009	0.014	0.007	0.032	0.018	0.067	0.008	0.244
134	0.020	0.042	0.028	0.009	0.014	0.007	0.032	0.018	0.011	0.064	0.244

FIGURE B5: Time component of credibility matrix Z^t ($Z_T^t \times multiplier$) at $t = 1$. (Color online)

https://doi.org/10.1017/asb.2017.31 Published online by Cambridge University Press

Node	Matrix entry for node										Row total
	111	112	121	122	123	124	131	132	133	134	
111	0.068	0.097	0.046	0.013	0.107	0.024	0.035	0.009	0.004	0.004	0.407
112	0.040	0.242	0.037	0.010	0.093	0.021	0.030	0.007	0.003	0.003	0.487
121	0.009	0.018	0.473	0.024	0.154	0.037	0.015	0.003	0.001	0.001	0.736
122	0.011	0.022	0.106	0.336	0.171	0.044	0.017	0.004	0.002	0.001	0.713
123	0.001	0.003	0.009	0.002	0.923	0.006	0.003	0.001	0.000	0.000	0.949
124	0.005	0.010	0.037	0.010	0.103	0.642	0.010	0.002	0.001	0.001	0.820
131	0.004	0.008	0.009	0.002	0.029	0.006	0.764	0.013	0.005	0.005	0.845
132	0.008	0.016	0.016	0.004	0.048	0.010	0.107	0.510	0.012	0.011	0.742
133	0.011	0.022	0.020	0.005	0.057	0.012	0.134	0.036	0.380	0.016	0.692
134	0.010	0.020	0.017	0.005	0.049	0.010	0.117	0.032	0.016	0.456	0.732

FIGURE B6: Total credibility matrix Z^t at $t = 2$. (Color online)

Node	Matrix entry for node										Row total
	111	112	121	122	123	124	131	132	133	134	
111	0.088	0.127	0.054	0.014	0.101	0.029	0.034	0.010	0.006	0.005	0.469
112	0.044	0.289	0.042	0.011	0.088	0.024	0.029	0.008	0.005	0.004	0.543
121	0.011	0.024	0.441	0.023	0.171	0.045	0.016	0.004	0.002	0.002	0.739
122	0.013	0.028	0.105	0.309	0.187	0.051	0.017	0.004	0.003	0.002	0.719
123	0.001	0.001	0.005	0.001	0.961	0.003	0.001	0.000	0.000	0.000	0.975
124	0.005	0.013	0.043	0.011	0.109	0.631	0.010	0.002	0.001	0.001	0.828
131	0.005	0.011	0.011	0.003	0.031	0.007	0.737	0.017	0.009	0.008	0.838
132	0.009	0.022	0.019	0.005	0.050	0.012	0.119	0.470	0.019	0.015	0.741
133	0.010	0.023	0.019	0.005	0.047	0.012	0.118	0.034	0.472	0.017	0.757
134	0.010	0.022	0.018	0.005	0.046	0.011	0.113	0.032	0.020	0.486	0.761

FIGURE B7: Total credibility matrix Z^t at $t = 3$. (Color online)