BAYESIAN INFERENCE FOR PARTIAL ORDERS FROM RANDOM LINEAR EXTENSIONS: POWER RELATIONS FROM 12TH CENTURY ROYAL ACTA

BY GEOFF K. NICHOLLS[1,a] JEONG EUN LEE[2,b] NICHOLAS KARN[3,c] DAVID JOHNSON [4,d] RUKUANG HUANG [5,e] AND ALEXIS MUIR-WATT [6,f]

[1]*Department of Statistics, The University of Oxford, UK.,* [a]*nicholls@stats.ox.ac.uk*

[2]*Department of Statistics, The University of Auckland, New Zealand ,* [b]*kate.lee@auckland.ac.nz*

[3]*Faculty of Arts and Humanities, University of Southampton, UK,* [c]*N.E.Karn@soton.ac.uk*

[4]*St Peter's college, The University of Oxford, UK ,* [d]*david.johnson@spc.ox.ac.uk*

[5]*Department of Psychiatry, University of Oxford, UK ,* [e]*rukuang.huang@jesus.ox.ac.uk*

[6]*Private researcher, London, UK ,* [f]*alexis.muirwatt@gmail.com*

In the eleventh and twelfth centuries in England, Wales and Normandy, Royal Acta were legal documents in which witnesses were listed in order of social status. Any bishops present were listed as a group. For our purposes, each witness-list is an ordered permutation of bishop names with a known date or date-range. Changes over time in the order bishops are listed may reflect changes in their authority. Historians would like to detect and quantify these changes. There is no reason to assume that the underlying social order which constrains bishop-order within lists is a complete order. We therefore model the evolving social order as an evolving partial ordered set or *poset*.

We construct a Hidden Markov Model for these data. The hidden state is an evolving poset (the evolving social hierarchy) and the emitted data are random total orders (dated lists) respecting the poset present at the time the order was observed. This generalises existing models for rank-order data such as Mallows and Plackett-Luce. We account for noise via a random "queue-jumping" process. Our latent-variable prior for the random process of posets is marginally consistent. A parameter controls poset depth and actor-covariates inform the position of actors in the hierarchy. We fit the model, estimate posets and find evidence for changes in status over time. We interpret our results in terms of court politics. Simpler models, based on Bucket Orders and vertex-series-parallel orders, are rejected. We compare our results with a time-series extension of the Plackett-Luce model. Our software is publicly available.

1. Introduction.

In rank-order data we are presented with a collection of lists ranking a common set of items from best to worst or first to last. A list might order items according to the preferences of an assessor, or the outcome of a multiplayer game, and may rank all elements in the set or just some subset presented to an assessor.

In this paper we analyse a time series of 371 lists recording the order in which 67 different bishops are named as witnesses to legal documents called *Royal Acta*. The data are available online (Sharpe et al., 2014). These documents date from the eleventh and twelfth century (see Supp A.2 for example lists). Just a small subset of the bishops are named in any given list but each bishop is present in many lists. In a list the bishops' names were written down by a clerk in an order that is known to reflect status (henceforth status in the context of witnessing, which might differ from status in other social contexts). The status of a bishop was partly determined by seniority and diocese. The first canon of the Council of London
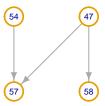

FIG 1. *A poset represented as a transitively closed DAG. In Sec. 7 we find evidence that this suborder contains all precedence relations between (54) Algar, Bishop of Coutances, (57) Adelulf, Bishop of Carlisle, (47) Simon, Bishop of Worcester and (58) Richard, de Beaufeu, Bishop of Avranches in 1136 (numbering as Fig 24).*

in 1075 concerns ecclesiastical precedence: "...each man shall sit according to his date of ordination, except for those who have more honourable seats by ancient custom or by the privileges of their churches" (Clover and Gibson, 1979). However, political standing may have contributed to status, and if it did, then the position of a bishop in the status-hierarchy would not be fully explained by time-in-office and diocese. Russell (1937) writes "The names of Eustace, bishop of Ely, and John, bishop of Norwich, frequently appear at the head of the list of bishops, before the names of the bishops of London and Winchester and of bishops who were consecrated before them... [however] ...they were very close friends of the king during most of his reign and were frequently at court". The dioceses of London and Winchester were nominally above Ely and Norwich so here is a case where some political element seems to count for more than seniority or "honourable seats". This analysis was contested by Haskins (1938) wrote "the precedence of witnesses to private grants is too erratic to serve as an index to their respective station". The question is still open. See Supp A.5 for a review of recent literature written by historians on this topic.

The set of precedence relations we reconstruct determine a *social hierarchy*. This follows the definition given in van Wietmarschen (2022), as the "socially expected behavior" for the clerk is to "value" by status. Precedence relations are transitive inequalities. This holds on social and historical grounds. Precedence is a social dominance relation and transitivity is fundamental to human and animal understanding of dominance (Gazes et al., 2017; Vasconcelos, 2008). Shizuka and McDonald (2012) find evidence for transitivity in 84 of 101 published animal dominance data tables, and it is the norm in models for social and organisational hierarchies (Friedell, 1967; Roberts, 1990), in early models for general preference relations (Bogart, 1973b) and in all analyses which assume an underlying complete order.

Gathering these observations, we represent the social hierarchy of bishops using a Directed Acyclic Graph (DAG). Nodes correspond to bishops and a directed edge indicates precedence. The DAG is transitively closed, so it defines a *partially ordered set* or "*poset*" (see Sec. 3.1.1, the terms are interchangeable). This is more general than assuming the social hierarchy is a total order, as would be the case if we fit a Mallows model. Some pairs of bishops may simply be unordered. The poset will be our parameter, the thing we want to estimate. Fig 1 gives an example of a poset extracted from a larger set of relations reconstructed in Sec. 6.3. The witness lists are our data. The poset constrains the data: bishop B shouldn't appear before bishop A in a witness list if B was ordered below A in the poset. In Sec. 3.1.3 we define an observation model which allows the lists to contradict the order relations in the reconstructed poset. However, we view this as as noise. The underlying structure of the social hierarchy of bishops in the context of witnessing is assumed to be a poset.

The first statistical work to model social networks using posets is Mogapi (2009) who writes "the application of partial orders in Social Network data has not been studied in the past". Martin (2002) used DAGs (which need not be transitively closed) in a similar setting without statistical inference. Friedell (1967) used semilattices, a subclass of posets in

which each pair of actors has a unique upper bound, to model organisational hierarchies. Their example, the "Cornerville S & A Club" hierarchy was a poset which only became a semilattice on dropping selected actors. Bogart (1973b) uses partial orders to express preference relations. They outline a model for evolving posets and remark "it would be extremely worthwhile to develop a theory of statistical inference for partial orderings". Fishburn and Gehrlein (1975) mention social dominance relations as a potential application in their work on posets. However, there seems to have been no statistical development of posets as models for social relations prior to Mogapi (2009) and no other precedent for posets to be used in the way we do. Some of our methods (excluding time series, covariates, and the theory in Sec. 4) were outlined in (Nicholls and Muir Watt, 2011) by two of the present authors and Muir Watt (2015) gives a continuous-time analysis without covariates. However, despite careful design of the particle filtering Monte Carlo, this approach does not seem promising from modelling and computational perspectives. Recent non-parametric work by Jiang and Nicholls (2024) shows how the dimension of the latent space parameterisation of posets may be estimated.

1.1. *Alternative Approaches.* Analysis of rank-order data (ie, our lists) often seeks a total order or actor-ranking which is "central" to the lists in the data, so that many lists are summarised by one complete actor ranking. Mallows models (Mallows, 1957) have a location parameter which is a ranking of the actor labels. In our setting this parameter would be interpreted as the unknown true bishop order. A dispersion parameter controls the distribution of the distance between realised lists and this centre-order. In Generalised Mallows models (Fligner and Verducci, 1986) the dispersion parameter can vary across rank positions. There is freedom in the choice of distance measure between ranking lists. Diaconis (1988) points to Kendall's-tau as having many good properties. Vitelli et al. (2018) adopt the foot-rule distance on modeling grounds and they and Irurozki et al. (2019) give methods and software (Irurozki et al., 2016; Sørensen et al., 2020) for computing the dispersion-dependent likelihood normalisation. Bayesian methods which allow for variability in the quality of the assessors providing the rankings (Deng et al., 2014) have been given. Mixture-model analysis (Meilă and Chen, 2010; Tkachenko and Lauw, 2016; Vitelli et al., 2018; Lu and Boutilier, 2014) can be used when there is a latent group structure in the population from which the lists are drawn, Meilă and Chen (2016) gives Bayesian methods for analysing non-parametric mixtures of generalised Mallows models and Vitelli et al. (2018) gives unsupervised clustering and treats incomplete lists. Asfaw et al. (2017) give time-series Mallows models.

The actor skill-vector in Plackett-Luce models (Luce, 1959; Plackett, 1975), which we discuss further in Supp J, determines a rank-order parameter which plays a similar role to the Mallows location parameter. Dispersion in the observation space of lists is controlled by the scale of the skill scores. Hunter (2004) gave an EM algorithm (see Caron and Doucet (2012)) for parameter estimation and much useful background theory. Bayesian methods Guiver and Snelson (2009), mixture models (Mollica and Tardella, 2017, 2020), time-series models (Caron and Teh, 2012; Glickman and Hennessy, 2015) and non-parametric Dirichlet process mixture models for clustering (Caron and Teh, 2012; Caron et al., 2014) have been developed. Caron and Doucet (2012) give efficient Monte Carlo methods exploiting data augmentation and conjugate priors. The Contextual Repeated Selection (CRS) model (Seshadri et al., 2020) generalises Plackett-Luce to handle non-transitive relations. This generality comes at a price. If we had ten lists of length two in which bishop A precedes B and ten in which B precedes C then A probably precedes C when they appear in the same list in our poset-based analysis. In the simplest CRS model at least, the twenty lists would provide no information about the ordering of A and C in a list containing only A and C.

Our poset-based parameterisation has strengths and weaknesses when compared to these models. Among the weaknesses, our likelihood evaluation does not scale to data with ordered

lists of actors longer than about 20 (though scaling with the number of lists is linear). This is discussed further in Sec. 3.1.2. In response, Jiang et al. (2023) restrict fitted posets to be Vertex-Series-Parallel posets (VSP, Valdes (1978)). These orders and Bucket Orders (in which unordered groups of actors are arranged in a total order) admit likelihood evaluation at a cost which is linear in the list length. Our model comparisons in Sec. 7 generally favor posets over VSPs and Bucket Orders. In other respects, our approach is closest to Placket-Luce as our latent variables play a similar role its skill scores. Our poset-model inherits some of the strengths of Placket Luce: while some Mallows and all Plackett-Luce (and our poset) models handle top-k data straightforwardly (where the assessor just ranks their top k preferences, see Sec. 3.1.3), fitting subset-data (where assessors are presented with different subsets for ranking) is more challenging for Mallows models, as the missing ranks have to be treated as missing data (Vitelli et al., 2018). This is not necessary in poset or Plackett-Luce models. On the other hand, Plackett-Luce models are "context independent": the probability for actor A to be listed above actor B is independent of the presence or absence of any other actor. Mallows, CRS and our poset-based models are in general context dependent. This property of the observation model defined in Sec. 3.1.2 is discussed further in Supp B.1.

Why is a qualitatively new ranking model needed, given the extensive range of models in the literature? First, we saw above that we have reason to believe the underlying social hierarchy is a partial order; we cannot estimate it reliably without fitting a model in which the parameter is a partial order. The alternative models described above impose a total order structure on the hierarchy; this is not justified and not needed. Secondly, we divide rank-order analyses into two classes: those which aim to reconstruct an underlying true or "physical" order and those in which the fitted order is understood as a heuristic summary of the lists. We work in the former setting. However, the models cited in Sec. 1.1 can be adapted to make heuristic models for our data. In Supp J.1 we specify and fit a Plackett-Luce time-series model with covariates. Glickman and Hennessy (2015); McKeough and Glickman (2024) define a Plackett-Luce model with many of the same features. Some conclusions from our poset-analysis can be obtained by fitting this relatively simpler model. However, our poset-model is simply a better model for our data, in the sense of goodness of fit. In Supp J.2 we estimate the Expected Pointwise Log Posterior Predictive (ELPD) model-selection measure (Vehtari et al., 2017) using Leave-One-Out Cross Validation (LOOCV) for our model and a Plackett-Luce mixture model (Mollica and Tardella, 2017). Our model is preferred.

1.2. *Statistical work with Partial Orders.* The first statistical methods inferring partial orders from list data were given in Mannila and Meek (2000) and Gionis et al. (2006). They treat problems of seriation in archaeology and biochronology in palaeontology and work with the VSP and Bucket Order sub-classes of posets for rapid evaluation of the "noise free" likelihood (see Sec. 3.1.2). Mannila (2008) gives a Bayesian analysis for Bucket Orders. In other important early work Beerenwinkel et al. (2007) define maximum likelihood posets and give a Bayesian analysis in Sakoparnig and Beerenwinkel (2012). They fit a probabilistic graphical model in which genetic mutations accumulate in a total order constrained by a poset. In related work Froehlich et al. (2007) model signaling pathways for gene expression and fit their models using simulated annealing.

In these archaeological and genetic settings there is an unknown true underlying poset and the data are total orders respecting that poset so we have the same data type and similar inferential goals. Our new contributions are as follows. In Sec. 3.1.3 we give a generative model for lists which allows for noise in the realised lists. We idealise the list-observation model as a snapshot of a "queue". In Sec. 3.2, building on work by Winkler (1985), we give marginally consistent priors with a hyper-parameter controlling the "depth" of a random poset. Depth is a quantity of historical interest, so our prior should be non-informative with respect to depth.

This rules out the uniform prior over posets, taken by Sakoparnig and Beerenwinkel (2012), as it is strongly informative of depth (see Supp D). In Sec. 3.2.3 we bring covariates into our model. We have a "linear predictor" which determines an actor's position in the poset. Finally, our list-data are a time series, so the generative model in Sec. 3.3 is a Hidden Markov Model (HMM) with a latent process of posets and "emitted data" which are lists of actors respecting the poset at the time each list was formed.

Posets appear in a range of data-analytic settings. In Mogapi (2009) the data are edges in a directed graph. The edges are noisy observations of the relations in an underlying poset representing information flow in a company, and a prior controls the number of relations in the order, like our focus on prior depth. This prior is not marginally consistent. Gionis et al. (2006) encode list data as a precedence matrix giving the proportion of times any pair of items appear in a given order. A poset has a corresponding precedence matrix, estimated using random *linear extensions* (lists which are total orders respecting the poset). The "distance" between lists and a poset is the distance between their precedence matrices. The estimated poset is a Bucket Order minimising this distance. Arcagni et al. (2022) has poset data and a wider range of otherwise similar loss functions. They fit both posets and Bucket Orders.

In Rising (2021) the poset is a summary statistic, displaying order relations between parameter estimates. Posets are also used for structure discovery in Bayesian Networks (Niinimäki et al., 2016; Kangas et al., 2016), where Bucket Orders support evaluation of marginal likelihoods. The likelihood is written as a sum over total orders respecting a Bucket Order and samples are reweighted by the order-count of the poset in an importance-sampling setup. The same count appears in our likelihood and we evaluate it using the same *lecount()* package (Kangas et al., 2019). However, in our setting the poset is a parameter of interest, not a supporting structure in the computation.

1.3. *Contributions and plan.* Our main contribution is our analysis in Sec. 6.3 of the bishop-list data. Our reconstruction of the evolving social hierarchy in Sec. 6 is the first statistical analysis of this kind of data. We answer some longstanding questions. How important was seniority in determining precedent? Did court politics play a role? Our poset models are also new (extending Mannila (2008) and Sakoparnig and Beerenwinkel (2012) as detailed above). Although our methodology was motivated by one particular data set, we nevertheless propose our poset-based ranking models as potentially useful for the analysis of rank data more broadly, to stand alongside Mallows, Plackett-Luce and other models for rank data.

In Sec. 2 we describe the list data, their associated dates and a seniority covariate on the bishops which informs their position in the hierarchy. Sec. 3 begins by setting out notation and defining partial orders and how they constrain the lists we actually observe. We motivate and define the observation model for lists in Sec. 3.1 and then in Sec. 3.2 give the prior. This is where we define "status" and how status is mapped to preference. The generative model and posterior are given Sec. 3.3. In Sec. 4 we show that our priors are marginally consistent and have support on every poset. We give a brief outline of our MCMC and define some useful summary statistics in Sec. 5, relegating the detail to Supp E. In Sec. 6 we present the results of our Bayesian-MCMC analysis. We begin in Sec. 6.2 with a sanity-check: we drop an order constraint on the seniority covariate-effects which historians expect to hold and show the order is recovered. In Sec. 6.3 we present our main results with the constraint now imposed. Results are discussed from a historical perspective in Sec. 6.4. In Sec. 7 we make model comparisons with other methods (fitting VSP and Bucket-Order models). Further comparison with variants of Plackett-Luce are given in Supp J. These favor our poset-model, though VSP and Bucket-Order models do quite well. We summarise our contribution in Sec. 8 and point to future work. A supplement discusses data registration, properties of the observation model and prior, MCMC, further results, model comparisons and results on synthetic data.

2. Introduction to the data.

2.1. *Context.* This study draws on an accumulated dataset, accessed through the database made for 'The Charters of William II and Henry I' project by the late Professor Richard Sharpe and Dr Nicholas Karn (Sharpe et al., 2014). Some historical background on the data is given in Supp A.1. Each witness list in the data is an ordered list of names of individual witnesses taken from a single legal document or "act" (collectively "acta"). A typical example (with List id 2364) is given in Supp A.2.

We have 1610 witness lists dated between 1066 CE and about 1166 CE involving 1760 individuals. A witness list is a "snapshot" created at a single event on a single day. We assume distinct lists are generated independently (for example, we see no evidence for a pair of Acta created at the same time with identical lists). Acta are witnessed in order of social rank from the king or queen, archbishop, bishops (as a group), earls (as a group, may precede bishops) and so on down through society. Historians ask if the order in which bishops appear within the their sub-list reflects their evolving personal authority. As we focus on the bishop-hierarchy we extract from the data the sub-lists of bishops. Many of the resulting lists contain less than two bishops and these are discarded as not informing relations between bishops. Data processing is set out in detail in Supp A.2.

We take time as discrete by year as the data gives dates rounded to the year. Further coarsening would mask recoverable structural change. Outside the range $[B = 1080, E = 1155]$ CE the lists are sparse so we focus on this interval, covering the reigns of William II, Henry I and Stephen and $T = E - B + 1 = 76$ years. The period is long enough for us to witness changes in the status of individual bishops, but short enough for there to be some hope of temporal homogeneity in the social conventions mapping status to witness list.

Data registration is detailed in Supp A.4. It leaves us with $N = 371$ lists, dated between 1080 and 1155, and containing two or more bishops. We refer to this as the "full data" (we look at shorter time intervals in our goodness-of-fit). Each of the $M = 67$ bishops in at least two lists is assigned a numerical index from 1 to M in Fig 24 in Supp G.2.

2.2. *Data notation.* Let $\mathcal{I} = \{1, ..., N\}$ and $\mathcal{M} = \{1, ..., M\}$ be the sets of list and bishop labels respectively. Each list $y_i = (y_{i,1}, ..., y_{i,n_i})$, $i \in \mathcal{I}$ is an ordered list of n_i bishops, so that $y_i \subset \mathcal{M}$, with bishop-$y_{i,1}$ first in the list, $y_{i,2}$ second and so on. Not all bishops appear in every list so the list is conditioned on "attendance". Let $o_i = \{y_{i,1}, ..., y_{i,n_i}\}$ be the *unordered* list of bishops in list i. Let $y = \{y_1, ..., y_N\}$ and $o = \{o_1, ..., o_N\}$.

We suppose the lists were generated by a process running over an interval of time $[B, E] = \{B, B+1, ..., E\}$. For $i \in \mathcal{I}$ let $\tau_i \in [B, E]$ give the time at which list i was created. This is sometimes uncertain. However, bounds $\tau_i^- \leq \tau_i \leq \tau_i^+$ are available. We estimate the missing list dates, taking a uniform prior on $\tau_i \in [\tau_i^-, \tau_i^+]$. Let $\tau = (\tau_1, ..., \tau_N)$, $\tau_i^\pm = (\tau_i^-, \tau_i^+)$ and $\tau^\pm = (\tau_1^\pm, ..., \tau_N^\pm)$.

Bishops enter the social hierarchy when consecrated and leave when they die. For $j \in \mathcal{M}$ we have (FEA, 2022) dates of consecration $b_j < E$ and death $e_j > B$ for each bishop. The distribution of these intervals can be seen in Supp A in Fig 16 at left. The intervals match the list date-ranges (τ_i^-, τ_i^+) so that no bishop appears in a list when not in post. At any given time some dioceses may be empty. Fig 2 shows the presence and absence of bishops by diocese. For $t \in [B, E]$ let $\mathcal{M}_t = \{j \in \mathcal{M} : b_j \leq t \leq e_j\}$ give the set of bishops active at time t, let $m_t = |\mathcal{M}_t|$ give the number of active bishops at time t and let

$$D = \max_{t \in [B,E]} m_t \tag{1}$$

give the greatest number active at any time ($D = 22$, in 1133). This quantity plays a role in bounding the required dimension of the model parameter space.

FIG 2. *Left axis gives dioceses. Right axis gives the number of lists in which each diocese appears. The x-axis gives the date in years. In each year a diocese may have a bishop in-post (dark cell) or be unoccupied (light).*

2.3. *Witness list data.* Bishops from thirty one dioceses appear in the data (including one, Tusculum, from Italy, dropped from the data). They are listed in Supp A.3 and can be seen at the left side of Fig 2. The dates of $|\{i \in \mathcal{I} : \tau_i^+ - \tau_i^- > 1\}| = 212$ lists are uncertain (the mean interval length is 4 years, and 90% span less than 10 years). Date intervals $[\tau_i^-, \tau_i^+]$ are plotted in Supp A in Fig 16 at right. Fig 3 plots lists and their lengths against their dates (using the midpoint of $[\tau_i^-, \tau_i^+]$, $i \in \mathcal{I}$). We include a list if at least half its interval falls within the 76-year interval $[B, E]$; most of the lists in our analysis fall entirely within it.

Our information about a bishop's status is limited by the number of lists in which a bishop appears. Longer lists are more informative as they inform relations between many pairs of bishops. Fig 4 shows the distribution of the number of lists a bishop appears in and the distribution of list lengths. Most bishops appear in a small number of lists and most lists are relatively short. However, lists "link together". If two bishops j_1, j_2 do not appear in any list together, but j_1 comes before j_3 in some list and j_3 before j_2 in another, then this is evidence for j_1 having higher status than j_2. This evidence accumulates over lists.
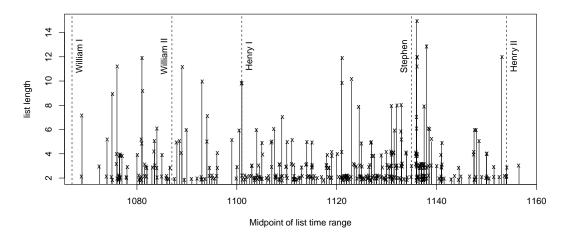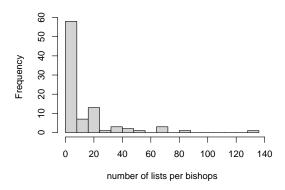

FIG 3. *List lengths and dates. Dashed vertical lines are coronation dates of Kings, bar heights are longest lists at that date, with a (jittered) cross for each list plotted at (date,length).*
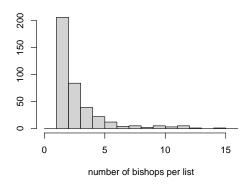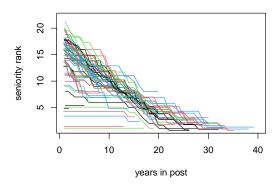
FIG 4. *Frequency of lists per bishop, a histogram of the counts $\sum_{i\in\mathcal{I}}\mathbb{I}_{j\in o_i}$, is plotted at left. Distribution of list lengths, a histogram of the counts $n_i = |o_i|$, is plotted at right.*
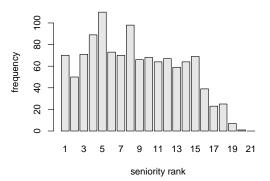

FIG 5. *Seniority rank covariate s defined in Eqn 2. Rank $s_{t,j}$ is plotted against "years in post", $t - b_j$, for each bishop $j \in \mathcal{M}$ (Left). The frequency of seniority-rank r in lists is plotted against r from 1 to S (Right).*

2.4. *Seniority covariates.* A bishop's "seniority" may have contributed to their overall status (Clover and Gibson, 1979) so it is a covariate in our model for the hierarchy of bishops. In year t the longest serving bishop has seniority-rank one and the last appointed bishop has rank at most m_t. Denote by $s_{t,j} \in \{1, ..., m_t\}$ the seniority-rank of bishop $j \in \mathcal{M}_t$ in year $t \in \{b_j, b_j + 1, ..., e_j\}$. We define seniority-rank as

$$(2) \qquad s_{t,j} = \sum_{k \in \mathcal{M}_t} \mathbb{I}_{b_j \geq b_k}.$$

Bishops have equal seniority if there are ties in the start dates b_j, $j \in \mathcal{M}$. The greatest seniority observed, $S = \max_{t,j} s_{t,j}$, is less than or equal D, the most bishops active. This is a second quantity which informs the dimension of the model we fit.

Fig 5 shows (at left) seniority-rank traces for each bishop from their first to last year in post. Bishops progress in rank by about one place every one or two years, more rapidly at first, as there are more bishops ahead of them. Our estimates of the effect of possessing seniority-rank $r \in \{1, ..., S\}$ depends for precision on a bishop with seniority r appearing in a reasonable number of lists, so we plot (Fig 5, right) the occurrence frequency $f_r = \sum_{i\in\mathcal{I}} \sum_{j\in o_i} \mathbb{I}_{s_{\tau_i,j}=r}$ against r to see which levels of the covariate are well represented in the data. Some dioceses were more peaceful and wealthy than others, so we considered taking diocese label as a second covariate for status. However, diocese would be colinear with bishop label, as each bishop only occupies one diocese in the period of study. An effect due to diocese would not be identifiable with the effects due to the bishops in that diocese.

2.5. *Key questions.* The key question is whether the status of bishop $j \in \mathcal{M}$ in year $t \in [B, E]$ was determined by their seniority $s_{t,j}$ in that year and their diocese or whether the personal authority or position in court of bishop j played a role. In terms of the list data y, are the positions of bishops in a list determined by seniority and diocese, with any variation away from this order just unstructured noise, or are the variations away from precedence rules structured by some kind of latent authority? The status of a bishop determines their place in the hierarchy, so what did the hierarchy look like in any given year? We answer these questions by building a statistical model for the bishop-list data.

3. Models and Inference. Our model, in which lists gathered in year t respect a social hierarchy which is known and respected by all but subject to occasional change, expresses the evolving social hierarchy. We present our model as a description of relations between actors, in the usual terminology of social network analysis. However, it can be applied to the analysis of any ranking list data, with or without time-series structure: Jiang et al. (2023) shows that a related fixed-time model is a good fit for Formula 1 race outcomes.

3.1. *Parameters and observation model.*

3.1.1. *Partial orders and linear extensions.* In this section we define precedence relations using posets. We drop the time dependence and consider a single generic observation. Suppose we have m actors with labels in $[m] = \{1, ... m\}$. We represent the unknown true order relations between actors as a poset on $[m]$. Brightwell (1993) gives an overview of models for random posets and is the source for much of what follows. A strong partial order $\succ_H$ on the ground set $[m]$ is a set of acyclic, transitively closed relations $i \succ_H j$ on the elements of $i, j \in [m]$. Ties $i \sim_H j$ are excluded. The relations in H are transitively closed if $i \succ_H j$ and $j \succ_H k$ implies $i \succ_H k$. The order is only partial as some elements are not ordered. A poset is a total order if $i \succ_H j$ or $j \succ_H i$ for every pair $i, j \in [m]$.

Partial orders on $[m]$ are one to one with transitively closed directed acyclic graphs (DAGs) with vertex labels $1, ..., m$, one vertex for each of the m actors, so $\succ_H$ is represented by a DAG $(H, [m])$ with edge set

$$H = \{\langle i, j \rangle \in [m] \times [m] : i \succ_H j\}.$$

See the example in Fig 6. We refer to transitively closed DAGs as if they were posets, as they correspond one to one. We can identify a poset by its edge set H as the edge set will be random while the vertex labels $[m]$ which define the ground set are always fixed. Since posets are edge sets we can take intersections of posets. This gives the poset with all relations shared by the intersected posets. The *dimension* of a poset $H \in \mathcal{H}_{[m]}$ is the smallest number of total orders which intersect to give H.

Let $\mathcal{H}_{[m]}$ be the set of all transitively closed DAGs on $[m]$ and let $H \in \mathcal{H}_{[m]}$ be a generic poset. For plotting purposes the transitive reduction is convenient. This is the unique DAG obtained from H by removing all edges implied by transitivity. The depth of a social hierarchy is of interest in many applications. The depth $d(H)$ of $\succ_H$ is the length of the longest path on the DAG H, so $d : \mathcal{H}_{[m]} \to [m]$. The poset in Fig 6 has $d(H) = 4$.

Let $\mathcal{P}_{[m]}$ be the set of all permutations of $[m]$. A *linear extension* of H is any list $\ell = (\ell_1, ... \ell_m)$, $\ell \in \mathcal{P}_{[m]}$ in which lesser entries come after greater entries, so $\ell_j \succ_H \ell_k$ is not allowed if $k < j$. For example, if H is the poset displayed in Fig 6 then H has three linear extensions, $(1, 2, 3, 4, 5), (1, 2, 4, 3, 5)$ and $(1, 4, 2, 3, 5)$. Denote by

$$\mathcal{L}[H] = \{\ell \in \mathcal{P}_{[m]} : \langle \ell_j, \ell_k \rangle \notin H \text{ for all } 1 \leq k < j \leq m\}$$

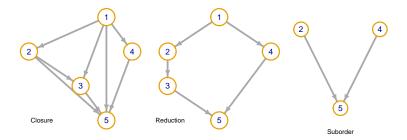the set of all linear extensions of $H \in \mathcal{H}_{[m]}$.

FIG 6. *Poset* $H = \{\langle 1,2 \rangle, \langle 1,3 \rangle, \langle 1,4 \rangle, \langle 1,5 \rangle, \langle 2,3 \rangle, \langle 2,5 \rangle, \langle 3,5 \rangle, \langle 4,5 \rangle\}$, $H \in \mathcal{H}_{[m]}$ *on* $[m] = (1,2,3,4,5)$ *represented by its transitively closed directed acyclic graph (left) and transitive reduction (centre). Suborder* $H[O]$ *for* $O = \{2,4,5\}$ *(right). The longest path on* H *is* $1 \to 2 \to 3 \to 5$ *so the depth is* $d(H) = 4$.

If we start with a social hierarchy $H \in \mathcal{H}_{[m]}$ over all actors then the hierarchy constraining any given subset $O \subseteq [m]$ of the actors is the *suborder*

$$\text{(3)} \qquad\qquad H[O] = \{\langle j_1, j_2 \rangle \in H : \{j_1, j_2\} \subseteq O\}$$

obtained by retaining edges between vertices in O. If H is a poset then so is $H[O]$. For example, in Fig 6, if $O = \{2,4,5\}$ then suborder $H[O]$ is the three-vertex DAG at right.

3.1.2. *Lists as randomly ordered queues.* A single generic list $Y = (Y_1, ..., Y_m)$ is modeled as a random linear extension of H. Let $C(H) = |\mathcal{L}[H]|$ be the number of linear extensions of poset $H \in \mathcal{H}_{[m]}$. The "noise free" likelihood for H is simply

$$\text{(4)} \qquad\qquad p(Y|H) = C(H)^{-1} \mathbb{I}_{Y \in \mathcal{L}[H]}.$$

All lists which respect the social hierarchy are equally likely. This "context dependent" observation model (see Supp B.1) is motivated by thinking of each list as a realisation of a random queue process in which actors not ordered by H randomly swap places. The equilibrium of this process is the uniform distribution on linear extensions of H (Karzanov and Khachiyan, 1991; Jiang and Nicholls, 2024), so if Y is a snapshot of this queue at equilibrium then $Y \sim \text{Unif}(\mathcal{L}[H])$. Witness lists were written down by a royal scribe with (we assume) perfect knowledge of the hierarchy H, so the queue model is an idealisation. However, we arrive at the same model if we assume the clerks regarded all orders not conflicting the hierarchy as equally likely.

Computation of $C(H)$ is #P-complete (Brightwell and Winkler, 1991) so no polynomial time algorithm for computing $C(H)$ is available, or is likely to exist, and evaluating $p(Y|H)$ is prohibitive at large m. However, in our data m is small enough to allow likelihood evaluation in reasonable time. For $j \in [m]$ let $\mathcal{L}_j[H] = \{\ell \in \mathcal{L}[H] : \ell_1 = j\}$ be the set of linear extensions with j first in the list and let $C_j(H) = |\mathcal{L}_j[H]|$. Partitioning on the first entry,

$$C(H) = \sum_{j=1}^{m} C_j(H),$$

which Knuth and Szwarcfiter (1974) compute using the suborder recursion in Supp B.2.

3.1.3. *Queue-jumping observation model for suborders.* In this section we define the observation model we use in our analyses. We modify the likelihood in Eqn 4 in three ways: lists may be "noisy"; just a subset of actors are in any given list; lists are observed over time.

We allow for noise in the observation model by allowing individuals to "jump the queue". A list Y is formed by taking individuals from the top of a queue which continues to mix

rapidly, constrained by the suborder on those remaining. Before the j'th actor is chosen, there are $m - j + 1$ individuals (with labels $Y_{j:m}$) yet to be placed. With probability p the next actor (ie Y_j) is chosen at random, ignoring any order constraints. Otherwise, Y_j is chosen as the first actor in a random linear extension of the suborder $H[Y_{j:m}]$ for those remaining. The fraction of lists headed by Y_j is $C_{Y_j}(H[Y_{j:m}])/C(H[Y_{j:m}])$. Working from the top down,

$$p_{(D)}(Y|H, p) = \prod_{j=1}^{m} p_{(D)}(Y_j|H[Y_{j:m}], p)$$

$$(5) \qquad = \mathbb{I}_{Y \in \mathcal{P}_{[m]}} \prod_{j=1}^{m} \left(\frac{p}{m-j+1} + (1-p)\frac{C_{Y_j}(H[Y_{j:m}])}{C(H[Y_{j:m}])} \right).$$

Noise allows any list to appear with non-zero probability. This is a "repeated selection" model (Seshadri et al., 2020) in which the next actor is chosen sequentially from those that remain. It follows that the correct likelihood for top-k data (where an assessor sees all m actors but just ranks their top $k < m$ actors) simply stops the product in Eqn 5 at $j = k$.

We take $p \sim \text{Beta}(1, \delta)$ as our family of priors for the queue-jumping probability p. The prior hyperparameter $\delta \geq 1$ is fixed (for example, in Sec. 6 we take $\delta = 9$, so the prior probability for a queue-jumping event is about ten percent), expressing the belief that if order relations are present then they are respected.

In Eqn 5 individuals are promoted up the queue. We can also model random "demotion". In this case the list is filled from the bottom up: with probability p the next actor is chosen at random, ignoring any order constraints; otherwise, they are the last entry in a random linear extension of the suborder for the remaining individuals. The likelihood becomes

$$(6) \qquad p_{(U)}(Y|H, p) = \mathbb{I}_{Y \in \mathcal{P}_{[m]}} \prod_{j=1}^{m} \left(\frac{p}{j} + (1-p)\frac{\tilde{C}_{Y_j}(H[Y_{1:j}])}{C(H[Y_{1:j}])} \right),$$

where $\tilde{C}_{Y_j}(H)$ is the number of linear extensions of H which end with Y_j. Jiang et al. (2023) extend these models to allow random promotion and demotion in a single realised list. The likelihood is tractable, but evaluation is time consuming so we do not fit that model here.

The noise free case is obtained from both Eqns 5 and 6 at $p = 0$. One consequence is that the noise-free model is also a repeated selection model. See Supp B.3 for the proof.

PROPOSITION 1. The noise-free and noisy models (Eqns 4, 5 and 6) all coincide at $p = 0$.

We now consider what happens when just a subset of actors are present. Relations are given by their suborder and the observation model applies for lists realised on suborders. Suppose that, when Y was realised, a subset $O = \{O_1, ..., O_n\}$, $O \subseteq [m]$ of actors entered the queue. Since they were constrained by the suborder $H[O]$, the noise free observation model is $Y \sim \text{Unif}(\mathcal{L}[H[O]])$: the list is a random draw from the linear extensions of the suborder. For example, for H in Fig 6, if $O = \{2, 4, 5\}$, the linear extensions are $\mathcal{L}[H[O]] = \{(2, 4, 5), (4, 2, 5)\}$ and Y is chosen at random from this set. The queue-jumping likelihoods $p_{(D)}, p_{(U)}$ are obtained by replacing $H \rightarrow H[O]$, $[m] \rightarrow O$ and $m \rightarrow n$ in Eqns 5 and 6.

3.1.4. *Time series of lists.* Finally, we restore time and give the full likelihood. This observation model is illustrated in Fig 7. At each time $t \in [B, E]$ the actors indexed in $\mathcal{M}_t$ were active and had precedence relations given by some poset $h^{(t)} \in \mathcal{H}_{\mathcal{M}_t}$. The sequence of partial orders from B to E is

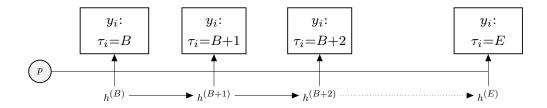$$(7) \qquad h = (h^{(B)}, h^{(B+1)}, ..., h^{(E)}),$$

FIG 7. *The observation model. The distribution of lists $\{y_i : i \in \mathcal{I}, \ \tau_i = t\}$ observed at time t is parameterised by a poset $h^{(t)}$ and noise parameter p. A stochastic process realising $h^{(t)}$, $t \in [B, E]$ is given in Sec. 3.2.*

where $h \in \mathcal{H}^{(B,E)}$ with

$$
\mathcal{H}^{(B,E)} = \mathcal{H}_{\mathcal{M}_B} \times \mathcal{H}_{\mathcal{M}_{B+1}} \times ... \times \mathcal{H}_{\mathcal{M}_E}. \tag{8}
$$

The model for h in Sec. 3.2 will define a poset process with poset time-series realisations.

For $i \in \mathcal{I}$, list y_i was formed under constraints imposed by the suborder $h^{(\tau_i)}[o_i]$, so in the noise-free model $y_i \in \mathcal{L}[h^{(\tau_i)}[o_i]]$. Allowing for noise, the likelihood is

$$
p(y|h, \tau, p) = \prod_{i=1}^{N} p(y_i | h^{(\tau_i)}[o_i], p), \tag{9}
$$

where $p(y_i | h^{(\tau_i)}[o_i], p)$ is given by $p_{(D)}$ or $p_{(U)}$ in Eqns 5 or 6 (depending on our choice of model) with the replacements $Y \to y_i$, $m \to n_i$ and $H \to h^{(\tau_i)}[o_i]$.

3.2. *Latent variables and covariates in a prior for partial orders.* We derive prior models for posets from k-dimensional random orders (Winkler, 1985). We modify this setup as we would like to have some control over the prior depth distribution. The depth of the social hierarchy of bishops is meaningful to historians so a prior which is non-informative with respect to depth will be useful. The uniform prior on posets certainly wouldn't be appropriate as it is strongly informative of depth: it concentrates on posets of depth three as $m \to \infty$ (Kleitman and Rothschild, 1975). The effect is illustrated in Supp D.

3.2.1. *Latent variable parameterisation.* In this section we define status and how it is mapped to precedence order relations. We use feature vectors, one for each actor, to determine actor-placing in the social hierarchy. See Fig 8 for illustration. These "status-features" do not correspond to any identifiable physical attributes. Following Winkler (1985), we associate with each actor $j \in \mathcal{M}_t$, at a time $t \in [b_j, e_j]$, a $1 \times K$ latent vector $Z_j^{(t)} \in \mathbb{R}^K$, $Z_j^{(t)} = (Z_{j,1}^{(t)}, ..., Z_{j,K}^{(t)})$ of $K \geq 1$ status-features. Let $Z^{(t)} = (Z_j^{(t)})_{j \in \mathcal{M}_t}$ be an $m_t \times K$ status matrix, with one row for each actor active at time t and one column for each status feature $1, ..., K$.

The partial-order $h^{(t)}$ at time t is a function of $Z^{(t)}$. At time t actor $j \in \mathcal{M}_t$ is above actor $j' \in \mathcal{M}_t$ if *all* status variables of j are greater than those of j', that is, $h^{(t)} = h(Z^{(t)})$ with

$$
h(Z^{(t)}) = \{\langle i, j \rangle \in \mathcal{M}_t \times \mathcal{M}_t : Z_{i,k}^{(t)} > Z_{j,k}^{(t)} \text{ for all } k = 1, ..., K\}. \tag{10}
$$

The setup is illustrated in Fig 8 at centre, where the Z-matrix has $m_t = 5$ rows and $K = 4$ columns. The rows of $Z^{(t)}$ are "paths" in the space $[K] \times \mathbb{R}$: the relation $\langle j, j' \rangle \in h^{(t)}$ holds when the path $(k, Z_{j,k}^{(t)})_{k=1}^{K}$ lies above the path through $(k, Z_{j',k}^{(t)})_{k=1}^{K}$; if the paths cross then the actors are unordered. In Fig 8, the Z-path for actor 4 intersects the paths for 2 and 3 so 4 is unordered with respect to 2 and 3. The other paths do not intersect so 1, 2, 3, 5 have a
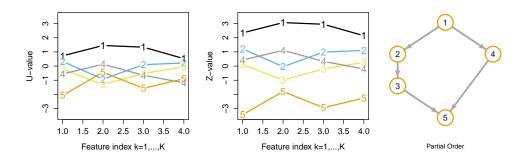
FIG 8. *Latent variable representation of the poset H in Fig 6: (left) U-matrix defined in Sec. 3.2.2; (centre) Z-matrix and (right) poset $h(Z)$ defined in Sec. 3.2.2. The paths in Z are shifted by the effects β.*

total order. This is a latent feature representation of the poset at right. In Winkler (1985) the columns of $Z^{(t)}$ are independent and paths are likely to cross. By contrast, the priors we give in Sec. 3.2.3 correlate columns and give us some control over the prior depth distribution.

Our mapping from $Z^{(t)}$ to $h^{(t)}$ is equivalent to intersecting the K total orders given by ranking the entries in each column of $Z^{(t)}$. See Supp C.2 for a discussion of this point. As we take a fixed value of K, the dimension of $h^{(t)}$ is at most K. However, using results from (Hiraguchi, 1951; Bogart, 1973a), it may be shown that if $K \geq \lfloor m_t/2 \rfloor$ then any poset $h^{(t)} \in \mathcal{M}_t$ can be represented by some $m_t \times K$ matrix $Z^{(t)}$. This is discussed in a modelling context in Muir Watt (2015) and proven in Sec. 4.2. Since $h^{(t)}$, $t \in [B, E]$ has at most D vertices (see Eqn 1), a model with $K \geq \lfloor D/2 \rfloor$ can represent any partial order process $h \in \mathcal{H}^{(B,E)}$.

Taking K as large as $\lfloor D/2 \rfloor$ is conservative as the true h-process may live in a lower dimensional space. In recent work Jiang and Nicholls (2024) estimates K, using a non-parametric prior for the U-process and sampling K using reversible jump MCMC. They considered smaller data sets with fewer lists all gathered at a single fixed time. In that setting the best fitting values of K were often much smaller than $\lfloor D/2 \rfloor$. In earlier work Durante et al. (2017b) used a similar approach to estimate the dimension of a latent space parameterisation of a population of networks. In both cases graph structures are modeled using latent continuous variables. Our approach is more like Durante et al. (2017a) and Gwee et al. (2023): the dimension of the latent space is chosen so that some kind of representation property holds (like our Proposition 3); shrinkage (Rousseau and Mengersen, 2011) effectively removes unwanted components. In our setting ρ controls this shrinkage. When it is larger the columns of Z tend to have the same order, so the dimension of the poset is smaller.

3.2.2. *Covariate effects for partial orders.* .

In Sec. 2.2 we introduced an actor-specific covariate informing status relations among bishops. In the following $s_{t,j} \in \{1, 2, \ldots, S\}$ is a single categorical or ordinal variable with levels from 1 to S. More general covariates are easily accommodated. Let $\beta \in \mathbb{R}^S$ be the vector of level effects. We split the *status*-vector $Z_j^{(t)}$ of actor $j \in \mathcal{M}_t$ into an additive effect $\beta_{s_{t,j}}$ due to $s_{t,j}$ and an *authority*-vector $U_j^{(t)} \in \mathbb{R}^K$, which captures all aspects of status which are not attributable to the covariate. Our additive model is, for $j \in \mathcal{M}_t$ and $t \in [B, E]$,

$$
(11) \qquad Z_j^{(t)} = U_j^{(t)} + 1_K \beta_{s_{t,j}}
$$

where 1_K is a row vector of K ones. Higher $\beta_{s_{t,j}}$-values lift all components of $U_j^{(t)}$ by a constant, raising the status features in $Z_j^{(t)}$. This moves the path $(k, Z_{j,k}^{(t)})_{k=1}^K$ above other paths and gives a higher position for actor j in the poset $h^{(t)}$. See Fig 8 for an example.

In our application the covariate is ordinal with a greater effect expected for lower values of s so we will be interested in testing for $\beta_1 > \beta_2 > ... > \beta_S$. Let $\mathcal{B}_0 = \mathbb{R}^S$ and

$$(12) \qquad \mathcal{B}_S = \{\beta \in \mathcal{B}_0 : \beta_1 > \beta_2 > ... > \beta_S\}.$$

We carry out analyses under models with $\beta \in \mathcal{B}_0$ (to check our prior expectation that $\beta \in \mathcal{B}_S$) and then again with $\beta \in \mathcal{B}_S$ (for best estimation with a well supported subjective prior).

Let $Z = (Z^{(t)})_{t=B}^E$ and $U = (U^{(t)})_{t=B}^E$ and write $Z = Z(U, \beta; s)$ for the function defined in Eqn 11. The parameters U and β replace h in the likelihood via $h = h(Z(U, \beta; s))$.

3.2.3. *Prior probability distributions.* We model the K-dimensional authority-process $U_j = (U_j^{(t)})_{t=b_j}^{e_j}$ for actor $j \in \mathcal{M}$ as a vector autoregression of order one with times-series correlation θ and covariance $\Sigma^{(\rho)}$. In our model latent authority-features are correlated from one time step to another with a drift towards zero. The setup is illustrated in Fig 9.

Our prior for the process U_j is independent over $j \in \mathcal{M}$ with correlation parameters $0 \leq \theta \leq 1$ and $0 \leq \rho \leq 1$. Let $\Sigma^{(\rho)}$ be a $K \times K$ covariance matrix with diagonal elements $\Sigma_{k,k}^{(\rho)} = 1$ and off diagonal $\Sigma_{k,k'}^{(\rho)} = \rho$ for $k, k' \in [K]$. Let 0_K be a vector of K zeros. For $j \in \mathcal{M}$ let

$$U_j^{(b_j)} \sim N\left(0_K, \frac{\Sigma^{(\rho)}}{(1-\theta)^2}\right)$$

and for $\epsilon_j^{(t)}$ iid for $t \in [b_j + 1, e_j]$ and each j,

$$(13) \qquad U_j^{(t)} = \theta U_j^{(t-1)} + \epsilon_j^{(t)}, \quad \epsilon_j^{(t)} \sim N(0, \Sigma^{(\rho)}).$$

Write $U_j \sim \text{VAR}_{K,\rho,\theta}^{(b_j,e_j)}(1)$, $j \in \mathcal{M}$ for the vector auto-regression with density

$$(14) \qquad \pi(U_j|\rho,\theta) = N\left(U_j^{(b_j)}; 0_K, \frac{\Sigma^{(\rho)}}{(1-\theta^2)}\right) \prod_{t=b_j+1}^{e_j} N\left(U_j^{(t)}; \theta U_j^{(t-1)}, \Sigma^{(\rho)}\right).$$

The parameter ρ controls the prior depth-distribution. When $\rho \simeq 1$ paths $(k, U_{j,k}^{(t)})_{k=1}^K$ are relatively flat as entries are strongly correlated. Flat paths don't intersect, so there are more order relations and $d(h^{(t)})$ is larger. When ρ is close to zero the paths are more jagged so there are few order relations and $d(h^{(t)})$ is smaller. We take as our prior $\rho \sim \text{Beta}(\gamma_1, \gamma_2, \gamma_3)$ with non-centrality parameter γ_3 and $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ fixed. Prior simulation (Supp D, Fig 17, left) showed $K = \lfloor D/2 \rfloor$ and $\gamma = (1, 1/3, 8)$ gave a prior on posets which is acceptably uniform on depth. Our prior on θ is uniform, $\theta \sim \text{Unif}(0, 1)$.

Our prior density for β is $\pi_\beta(\beta) = N(\beta; 0, I_S)$. The variation between levels of a covariate equals the variation in authority over one time step: $\beta_j - \beta_{j'}$ has the same variance as the components of $U_j^t - U_j^{(t-1)}$ in Eqn 13. We set $\Sigma_{k,k}^{(\rho)} = 1$ as we can scale ρ and the variance of β to get the same distribution for h. It is necessary to take proper priors for U and β. We discuss these priors further in Sec. 5.2 in relation to identifiability.

3.3. *Prior summary and Posterior distribution.* We now summarise our generative model for the data. The model is depicted in Figs 7 and 9. The data are the lists y. We condition on knowledge of the uncertainty ranges $\tau^\pm$, the covariate data s, the actor activities $\mathcal{M}_t$, $t \in [B, E]$ and the prior hyper-parameters $\gamma, \delta > 0$ and $K \geq 1$. The generative model is,

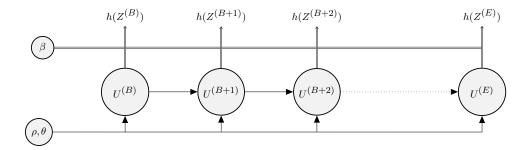$$\rho \sim \text{Beta}(\gamma), \qquad \text{with } \gamma = (1, 1/3, 8) \text{ unless stated,}$$

FIG 9. *Hidden layer of HMM: a latent process $U^{(t)}$ determines a sequence of posets $h^{(t)} = h(Z(U^{(t)}, \beta, s)$ for $t \in [B, E]$. The U-prior has hyperparameters ρ and θ, and double lines indicate implication $\Rightarrow$. This figure connects with Fig 7 at the top row of nodes $h^{(t)} = h(Z^{(t)})$ to give the generative model for the data.*

$$\theta \sim \mathrm{Unif}(0,1),$$

$$U_j \sim \mathrm{VAR}^{(b_j,e_j)}_{K,\rho,\theta}(1), \qquad \text{with } K = \lfloor D/2 \rfloor \text{ unless stated, and } U \text{ in Eqn 13,}$$

$$U_j^{(t)} \text{ iid for } j \in \mathcal{M} \text{ and defined for } t : j \in \mathcal{M}_t \text{ and}$$

$$\beta \sim N(0, I_S), \qquad \text{either } \beta \in \mathcal{B}_0 \text{ or constrained } \beta \in \mathcal{B}_S \text{ per Eqn 12.}$$

These collectively determine the partial-order prior via

$$Z = Z(U, \beta; s), \qquad \text{from Eqn 11, giving } Z = (Z^{(t)})^E_{t=B}, \text{ and}$$

(15) $$h = h(Z(U, \beta; s)), \qquad \text{from Eqn 10, giving } h = (h^{(t)})^E_{t=B};$$

Priors for the remaining observation model parameters are

$$\tau_i \sim U\{\tau_i^-, \tau_i^+\}, \qquad \text{independently for } i = 1, ..., N, \text{ and}$$

$$p \sim \mathrm{Beta}(1, \delta); \qquad \text{with } \delta \geq 1 \text{ and } \delta = 9 \text{ by default.}$$

Finally the data are realised

(16) $$y_i \sim p(\cdot | h^{(\tau_i)}[o_i], p), \qquad \text{independently for } i \in \mathcal{I},$$

using the distribution for y_i given in Eqn 5 or 6. The joint posterior distribution is

(17) $$\pi(\rho, \theta, U, \beta, \tau, p | y) \propto \pi(\rho, \theta, \beta, \tau, p) \pi(U | \rho, \theta) p(y | U, \beta, \tau, p),$$

where $p(y | U, \beta, \tau, p) = p(y | h(Z(U, \beta; s)), \tau, p)$ in Eqn 9 and $\pi(U | \rho, \theta) = \prod_{j \in \mathcal{M}} \pi(U_j | \rho, \theta)$ with $\pi(U_j | \rho, \theta)$ given in Eqn 14. The model without covariates or time is set out in Supp L.

4. Properties of priors. Our partial order prior is marginally consistent and expresses any partial-order time series in $\mathcal{H}^{(B,E)}$. Supp D explores the prior using simulation.

4.1. *Marginal consistency.* Marginal consistency is a relationship between members of a family of distributions. Dropping time, suppose that for each non-empty $O \subseteq [m]$ we write

down a prior $\pi_{\mathcal{H}_O}(G)$ on $G \in \mathcal{H}_O$. These priors are marginally consistent if for each O and $H \sim \pi_{\mathcal{H}_{[m]}}$ we have $H[O] \sim \pi_{\mathcal{H}_O}$ for the distribution of the suborder, that is,

$$(18) \qquad \pi_{\mathcal{H}_O}(G) = \sum_{H \in \mathcal{H}_{[m]}} \pi_{\mathcal{H}_{[m]}}(H) \mathbb{I}_{G=H[O]}.$$

The point here is that we define $\pi_{\mathcal{H}_O}$ for each $O \subseteq [m]$ and then verify Eqn 18. It can fail to hold if we write down a distribution $\pi_{\mathcal{H}_O}$ for each O without care. For example, the uniform distributions $\mathrm{Unif}(\mathcal{H}_O)$, $O \subseteq [m]$ are not marginally consistent. There are three partial orders for $m = 2$ and nineteen for $m = 3$, so if $H \sim \mathrm{Unif}(\mathcal{H}_{[3]})$ then we won't have $H[(1,2)] \sim \mathrm{Unif}(\mathcal{H}_{[2]})$ as we can't group nineteen posets into three equal-sized groups.

Winkler (1985) shows marginal consistency when the columns of $Z^{(t)}$ are independent. We extend this to our more general setting. We take $\beta = 0_S$, so no covariates, as we cannot expect marginal consistency when we have covariate information which explicitly breaks it. Let $\pi_{\mathcal{H}^{(B,E)}}(h|\beta = 0_S)$ be the marginal prior for $h \in \mathcal{H}^{(B,E)}$ when $\beta = 0_S$ (see Eqn 24 in Supp C.1). Let $\mathcal{H}^{(B,E)}_{-j}$ be the set of all partial order time-series with $j \in \mathcal{M}$ removed and for $h \in \mathcal{H}^{(B,E)}$ let h_{-j} be the suborder obtained by removing j.

PROPOSITION 2. Let $g \in \mathcal{H}^{(B,E)}_{-j}$ be given. Our priors are marginally consistent, that is

$$\pi_{\mathcal{H}^{(B,E)}_{-j}}(g|\beta = 0_S) = \sum_{h \in \mathcal{H}^{(B,E)}} \mathbb{I}_{g=h_{-j}} \pi_{\mathcal{H}^{(B,E)}}(h|\beta = 0_S).$$

for each $j \in \mathcal{M}$. [See Supp C.1 for proof]

Proposition 2 establishes marginal consistency for removing one element of $\mathcal{M}$. Consistency for more general marginals follows by removing elements one at a time. See Supp C.1 for remarks on settings where marginal consistency holds with covariate effects retained.

4.2. *Universal representation.* We claimed in Sec. 3.2.1 that, if $K \geq \lfloor D/2 \rfloor$ with D defined in Eqn 1 then any poset $h^{(t)} \in \mathcal{H}_{\mathcal{M}_t}$ can be represented by some $m_t \times K$ matrix $Z^{(t)}$. We restore $\beta \in \mathcal{B}_0$ or $\mathcal{B}_S$ and covariate effects. Let

$$(19) \qquad \pi_{\mathcal{H}^{(B,E)}}(h) = \int_{\mathbb{R}^S} \pi_{\mathcal{H}^{(B,E)}}(h|\beta)\pi(\beta)\,d\beta$$

be the full marginal prior with variable β, extended from Eqn 24.

PROPOSITION 3. Suppose $\min_{t \in [B,E]} m_t \geq 4$. The probability $\pi_{\mathcal{H}^{(B,E)}}(h)$ in Eqn 19, given by the generative model Eqn 15 with $K \geq \lfloor D/2 \rfloor$, assigns a positive probability mass $\pi_{\mathcal{H}^{(B,E)}}(h) > 0$ to every time-series $h \in \mathcal{H}^{(B,E)}$. [See Supp C.2 for proof.]

5. Computational methods.

5.1. *Markov Chain Monte Carlo.* We implemented an MCMC algorithm targeting $\pi(\rho, \theta, U, \beta, \tau, p|y)$ in Eqn 17. Each update is a simple Metropolis-Hastings MCMC step. The updates are summarised in Supp E. We tested the software evaluating the likelihood by simulating synthetic data and checking list proportions matched their probability in the likelihood. We also recover the true parameters of synthetic data (see Fig K.1). We run the MCMC producing L samples (after burn-in and thinning) $\rho^{(l)}, \theta^{(l)}, U^{(t,l)} = (U^{(t,l)}_j)_{j \in \mathcal{M}_t}, \beta^{(l)}, \tau^{(l)}$ and $p^{(l)}$ for $l = 1, ..., L$. This determines samples for $Z^{(t,l)} = (Z^{(t,l)}_j)_{j \in \mathcal{M}_t}$, with

$$Z^{(t,l)}_j = U^{(t,l)}_j + 1_K \beta^{(l)}_{s_{t,j}}$$

and similarly $h^{(t,l)} = h(Z^{(t,l)})$, for $t = B, ..., E$ and $l = 1, ..., L$.

The likelihood is not differentiable in U and β ruling out Hamiltonian MCMC. We tried updating sections of the time series in parallel (exploiting the Markov structure), halving the runtime at best. The bottleneck is ultimately the likelihood evaluation which is $\#P$-complete in list length. When lists are short the method scales well in the number of actors and the number of lists. The latent-parameter dimension $\dim(U) = K \times \sum_t m_t$ becomes the limiting factor (in Sec. 6.3, $\dim(U) = 13,453$). These challenges motivated work on VSP-models in (Jiang et al., 2023) which scale to lists with hundreds of actors.

5.2. *Posterior summaries.*   Besides plotting marginals for individual parameters ρ, θ, p and β, we report selected summary statistics computed on the MCMC output. These are the consensus poset (which displays relations with posterior support greater than one half, see Supp F) and the Bayes factor for the first S' of the S covariate effects to be ordered, $\mathcal{B}_{S'} = \{\beta \in \mathcal{B}_0 : \beta_1 > \beta_2 > ... > \beta_{S'}\}$ and given by $B_{S',0} = p(y|\beta \in \mathcal{B}_{S'})/p(y|\beta \in \mathcal{B}_0)$. Formulae for estimating these quantities are given in Supp F.

5.2.1. *Non-identifiability of authority and seniority effects.*   We are interested in separating the relative authority $U_j^{(t)}$ of a actor from the status $Z_j^{(t)}$. There are two sources of non-identifiability. The latent variables U have a label swapping symmetry: the posterior is invariant under permutation of the columns of $U^{(t)}$ if the same permutation is applied at each $t \in [B, E]$. One simple summary which is invariant under column permutation is the row-average. This gives the estimated average posterior authority for actor j at time t,

$$(20) \qquad \bar{U}_j^{(t)} = \frac{1}{L} \sum_{l=1}^{L} \frac{1}{K} \sum_{k=1}^{K} U_{j,k}^{(t,l)}.$$

A plot of $\bar{U}_j^{(t)}$ against t shows j's changing authority. Average status $\bar{Z}_j^{(t)}$ is defined similarly.

The second source of non-identifiability is shift invariance of $h^{(t)}$ under

$$(21) \qquad U_j^{(t)} \rightarrow U_j^{(t)} + 1_K u^{(t)}, \; t \in [B, E], \; j \in \mathcal{M}_t,$$

$$(22) \qquad \beta_r \rightarrow \beta_r + c, \; r \in [S],$$

where $u^{(t)} \in \mathbb{R}$ is a common shift applied to each actor, which may vary over t, and $c \in \mathbb{R}$ is common shift applied to all effects. The proper U and β priors shrink these shifts towards zero. We project these degrees of freedom out by subtracting the averages, $\bar{U}_j^{(t)} \rightarrow \bar{U}_j^{(t)} - \mathcal{M}_t^{-1} \sum_j \bar{U}_j^{(t)}$ and $\beta_r \rightarrow \beta_r - S^{-1} \sum_{r'} \beta_{r'}$, before computing the summary statistics and plotting. A similar issue arises in the Plackett-Luce time-series model in Supp J.1.

6. Results.

6.1. *Fitted models.*   Prior distributions are summarised in Sec. 3.3. Unless otherwise indicated we present results for the likelihood $p_{(D)}$ in Eqn 5 as clerks wrote lists from top down. Results are near-identical with $p_{(U)}$ in Eqn 6. Experiments showed fractionally lower estimated noise probabilities p for $p_{(U)}$ than $p_{(D)}$ (see Fig 21). This suggests the $p_{(U)}$ model is a slightly better fit, but there is little difference. The greatest number of active bishops is $D = 22$ so we take $K = 11$ (see end of Sec. 3.2.1) for the dimension of the latent status feature vectors $Z_j^{(t)}$, $j \in \mathcal{M}, t \in [B, E]$ in our main analyses in Secs. 6.2 and 6.3. We check robustness by taking $K = 2$ and $K = 22$ in Sec. G.3 of Supp G. The β-dimension is $S = 21$ (less than D as there is a tie at seniority-rank 21 in 1133, the year with the most active bishops).
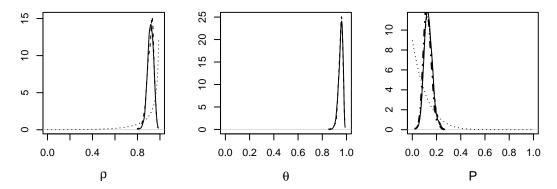
FIG 10. *Posterior densities for ρ, θ and p from the unconstrained seniority effects analysis in Sec. 6.2. Two independent MCMC runs are shown (solid and dashed). The dotted line in the ρ and p graphs is their prior. The prior for θ is uniform. The thick dash-dot curve in the p-density is the posterior density for a uniform prior on p.*
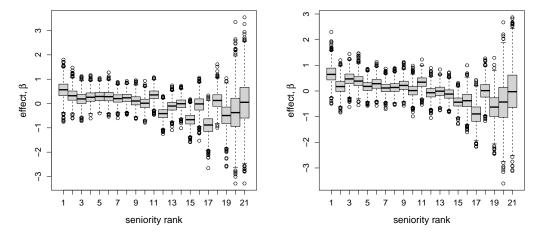

FIG 11. *(Left) Marginal posterior distributions of seniority effect parameters from the unconstrained seniority effects analysis in Sec. 6.2 with likelihood $p_{(D)}$ in Eqn 5. (Right) As left with likelihood $p_{(U)}$ in Eqn 6. See Fig 30 in Supp J.1 for the corresponding plot for the Plackett-Luce time-series model comparison.*

6.2. *Analysis with unconstrained seniority effects.* We begin by presenting our results for the full data set defined in Sec. 2.3. We first check that we see declining seniority effect at lower seniority, so we do not constrain the seniority effects to be ordered and take $\beta \in \mathcal{B}_0$.

Traces in Fig 22 in Supp G.1 for MCMC targeting the posterior in Sec. 3.3 show convergence. Marginal posterior densities from two independent runs are shown in Fig 10 and are near-identical. The correlation ρ of features in $U_j^{(t)}$, $j \in \mathcal{M}$ at each fixed time $t \in [b_j, e_j]$ is close to one, supporting relatively deeper posets. The time-series correlation parameter θ is close to one, indicating strong serial correlation between $U_j^{(t)}$ and $U_j^{(t+1)}$, and therefore also $h^{(t)}$ and $h^{(t+1)}$. Finally, the error probability p, which controls the extent to which lists y_i may depart from the linear extensions $\mathcal{L}[h^{(\tau_i)}]$, is small, as we would expect if the poset model captures the variation in lists. The prior for p has $\delta = 9$, favoring small p. We checked robustness to this choice: Fig 10 superimposes the posterior density when $\delta = 1$ (uniform, estimated by importance sampling); the shift from the $\delta = 9$ posterior is barely discernible.

In Fig 11 we plot marginal posterior β distributions in the posterior with likelihood $p_{(D)}$ (noise is random upward displacement) and $p_{(U)}$ (random downward displacement) respectively. There is a clear downward trend with increasing seniority-rank value (ie, lower senior-
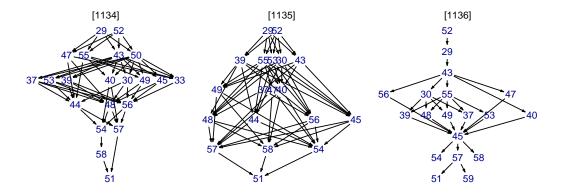
FIG 12. *(constrained seniority effects analysis of Sec. 6.3) Three of seventy-six years from one MCMC-sampled state, $h^{(t,L)}$, $t \in \{1134, 1135, 1136\}$. Vertex numbers correspond to bishops names in Fig 24 in Supp G.2.*

ity) as we expect. When the rank is large (18-21) we have few instances of bishops with that rank (see Fig 5), and distributions trend back to the prior. Under both observation models $p_{(U)}$ and $p_{(D)}$, the effect β_{11} is "out of order". This is because several bishops who spent several years at seniority-rank 11 (William Giffard, bishop of Winchester, Richard de Belmeis I, bishop of London and Henry de Blois, Bishop of Winchester) were connected with royalty, so it was their authority and not their seniority which pushed them up the lists. This is best modeled by imposing the seniority-effect order constraint $\beta \in \mathcal{B}_S$ as we do in the next section. In Sec. F.2 we test $\beta \in \mathcal{B}_S$ against the unconstrained model $\beta \in \mathcal{B}_0$ by estimating the Bayes factor, using a Savage-Dickey estimator. We find clear evidence in favor of the constraint. In summary we see in this first analysis the structures we anticipated.

6.3. *Analysis with constrained seniority effects.* The assumption of decreasing seniority effect with decreasing seniority rank is supported on historical and statistical grounds so we now impose the constraint $\beta \in \mathcal{B}_S$. We omit the ρ, θ and p posterior densities as they are essentially unchanged from Fig 10. Marginal posterior distributions for the unknown list dates τ_i, $i \in \mathcal{I}$ are given in Fig 26 in Supp G.2.

In Fig 12 we plot posets $h^{(t,L)}, t \in [1134, 1136]$ from one MCMC sample state (the final sample in the MCMC sample output $(h^{(t,l)})_{t \in [B,E]}$, $l \in [L]$). These three years bracket 1135 when Stephen became king. The number and length of lists in this period is relatively large (see Fig 3). In Fig 13 we plot posterior consensus posets $\bar{h}^{(t)}$ estimated at the same years (transitive reductions for ease of viewing, see Supp G.2 for all years). The transtive closures in Fig 23 in Supp G.2 have many more strongly supported edges as a chain of weakly supported relations give strongly supported relations from chain head to tail.

In Fig 25 in Supp G.2 we plot the evolving mean status values $\bar{Z}_j^{(t)}$ for each bishop as a function of time. Bishops are grouped by diocese. These curves have a "sawtooth" pattern, as the "status" measure Z trends up through the tenure of a bishop as their seniority increases. It drops down when a new bishop enters the diocese with low seniority. By contrast the curves in Fig 14 show the evolving mean authority values $\bar{U}_j^{(t)}$ for each bishop. These curves are flatter as the effect of seniority is removed. Nigel, bishop of Ely is revealing. His status in Fig 25 is fairly flat. This is because his mean authority $\bar{U}_j^{(t)}$ declined as his seniority increased.

The continuity in authority (but not status) of bishops over time within a diocese in Fig 14 is noteworthy. There are some exceptions. For example, Henry de Blois started with higher authority than might be expected based only on the diocese. Some dioceses seem to be better (Winchester, London, Lincoln) than others (Chichester, Rochester, the dioceses in Normandy). The bishops of London and Winchester had gained precedence over their colleagues
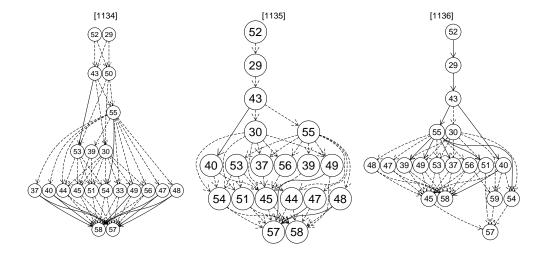
FIG 13. *(constrained-effects model Sec. 6.3) Consensus posets for 1134-1136 (transitive reductions). For closures see Supp G.2 Fig 23. Dashed/solid edges have posterior support greater than* $0.5/0.9$. *Numbering as Fig 12.*

at the Council of London in 1075 and Lincoln came in the later middle ages to rank after Winchester. However, there is uncertain evidence from as early as 1138 that the bishop of Lincoln might assume the role of London or Winchester in their absence and consequently that Lincoln already enjoyed a degree of precedence (Johnson, 2013).

6.4. *Discussion of results.* From a historical perspective, there are three significant outcomes. The first is the strong emphasis on the seniority and precedence of individual bishops in the witness lists. Historians often link the relative position of witnesses to an assessment of their political significance, but the analysis here shows that royal scribes were strongly influenced by the rules on seniority and precedence expressed at the Council of London, held by the English church in 1075 (Council of London, 1075, clause 1, Clover and Gibson (1979)).

The second is the position of Normandy within the Anglo-Norman realm. Fig 14 shows that early in this period (before about 1100) Norman bishoprics enjoyed high status, but that this declined from the early twelfth century. This change is particularly marked for Avranches, Bayeux and Évreux, whilst no English bishoprics show a comparable trend. This should inform the ongoing debate about the relationship between England and Normandy (Bates (2013), chapter 5). This change might represent a principled decision by royal scribes to rank Norman bishops lower in precedence than their English counterparts, or it might be explained politically. The smaller Norman dioceses may have been less attractive to ambitious clergymen, and there were periods when Normandy and England were ruled separately (most notably, 1144-54), so that Norman bishops were external to the English kingdom.

The third concerns how far the behaviour of individual bishops could change their status. Bishops were active politically and could fall into disgrace. Thus, Bishop Nigel of Ely had high status for a junior bishop in the 1130s, but from his disgrace in 1139 his status fell, contrary to the usual pattern. Nigel's pattern is unique; it is not replicated by that for other disgraced bishops, such as Ranulf Flambard of Durham after 1100 and Alexander of Lincoln after 1139. These differences presumably reflect the nature of the disgrace itself. The estimated poset relations accord with known political favour. For example, Henry of Blois and Odo of Bayeux, who were related to the royal house (Odo was William the Conqueror's half-brother, and Henry was Stephen's brother), are highly ranked. Referring to Fig 27 in
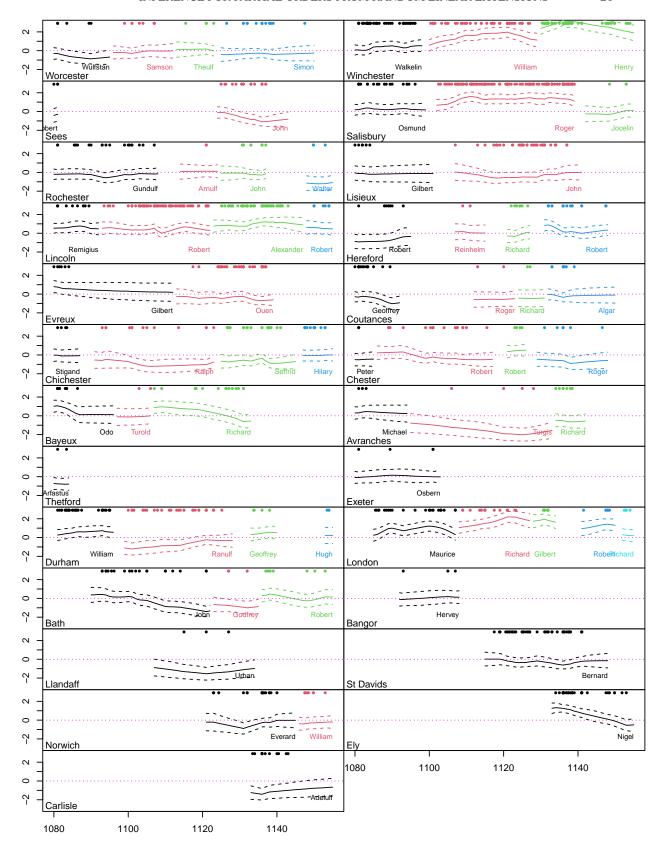
FIG 14. *(constrained seniority effects analysis of Sec. 6.3) Bishop-authority curves $\bar{U}_j^{(t)}$ (y-axis values, solid curves) plotted for each bishop $j \in \mathcal{M}$ as a function of time (x-axis, the year) from b_j to e_j with uncertainty at one-sigma. The dots at the top of each graph show the times of lists in which the matched bishop below appeared.*

Supp G.2, although Henry (52) did well from the start, until 1134 he shares top spot in consensus orders with Roger of Salisbury (29). From this date he is promoted ahead of anyone else. This suggests that his brother becoming king in 1135 had an impact on his position.

Referring to Fig 27, reconstructed orders seem relatively shallow, typically one (1097) or two (1124) groups of middle-ranked bishops and a few above or below. We may be concerned that this reflects an overly informative prior and differences in how often bishops witness. We tested this by simulating synthetic data in which the true posets were total orders (see Supp K.2), but with the same list memberships as the real data. We reconstructed the true total orders well. If the true orders were total orders we would see this in our analysis.

7. Comparisons with other models. In this section we define models over Bucket Orders and VSP orders. Calculation of $|\mathcal{L}[H]|$ for $H \in \mathcal{H}_\mathcal{A}$, $\mathcal{A} = \{1, \ldots, m\}$ is linear in m on these subspaces (Wells, 1971), so if these models were preferred then we would use them. We find they are not adequate to represent a time-evolving hierarchy over long periods of time but can give a good fit over short time periods. Jiang et al. (2023) applies a fixed-time VSP model to all the witness list data (not just bishops). Some orders have over 200 actors, with lists exceeding 50 in length, and are out of reach for our full poset analysis.

In Supp J.1 and J.2 we compare our model with Plackett-Luce models. Bayesian analysis of time-series Plackett-Luce in Supp J.1 gives similar results for a parameter function corresponding to the average authority in Eqn 20. Analysis of a Plackett-Luce mixture model in Supp J.2 on short time intervals shows our model is preferred.

7.1. *Bucket Orders and Vertex-Series-Parallel partial orders.*

VSP orders (Valdes, 1978; Valdes et al., 1982) are built recursively from the ground set by taking series and parallel combinations of posets. We give this intuitive definition in Supp I. Valdes et al. (1982) gives a concise characterisation. For any set $\mathcal{A}$, the class of all VSPs $\mathcal{V}_\mathcal{A}$ is identical to the set of posets $H \in \mathcal{H}_\mathcal{A}$ which do not contain a set of vertices $\mathcal{A}' = \{j_1, ..., j_4\}$ with sub-graph $H' = H \cap (\mathcal{A}' \times \mathcal{A}')$ isomorphic to the "forbidden subgraph" $F = \{\langle 1, 2 \rangle, \langle 3, 2 \rangle, \langle 3, 4 \rangle\}$ shown in Fig 29 at right. After vertex relabelling, F and H' must be identical, so edges absent in F are absent in H'. This makes it straightforward to test if a poset H is a VSP-order.

A sub-class of VSPs called "Bucket Orders" has a particularly simple closed form for $|\mathcal{L}[H]|$. Actors are grouped into "buckets". Actors in the same bucket are unordered and a complete order holds over buckets. Formally, if $\mathcal{K}_\mathcal{A}$ is the class of Bucket Orders on $\mathcal{A}$ then $b \in \mathcal{K}_\mathcal{A}$ iff there is a partition $\mathcal{A}_1, ..., \mathcal{A}_P$ of $\mathcal{A}$ into P buckets such that for each $k \in [P]$ and all $j_1, j_2 \in \mathcal{A}_k$ we have $\langle j_1, j_2 \rangle \notin b$ and for all pairs $1 \le k_1 < k_2 \le P$ of buckets and all $j_1 \in \mathcal{A}_{k_1}$ and $j_2 \in \mathcal{A}_{k_2}$ we have $\langle j_1, j_2 \rangle \in b$. VSPs and Bucket Orders are a small subset of partial orders. For example, if $|\mathcal{A}| = 18$ (the largest for which OEIS Foundation Inc (2022) gives cardinalities) we have $|\mathcal{H}_{[18]}| \simeq 2 \times 10^{35}$, $|\mathcal{V}_{[18]}|/|\mathcal{H}_{[18]} \simeq 10^{-11}$ and $|\mathcal{K}_{[18]}|/|\mathcal{H}_{[18]}| \simeq 10^{-17}$.

7.2. *Bucket and VSP-order models.*

Suppose we are interested in learning about order relations over a period $[t_1, t_2]$ with $B \le t_1 \le t_2 \le E$. If we could justify restricting the process of fitted posets $h \in \mathcal{H}^{(t_1, t_2)}$ to a VSP-order-process $h \in \mathcal{V}^{(t_1, t_2)}$ with

$$\mathcal{V}^{(t_1, t_2)} = \mathcal{V}_{\mathcal{M}_{t_1}} \times \mathcal{V}_{\mathcal{M}_{t_1+1}} \times ... \times \mathcal{V}_{\mathcal{M}_{t_2}},$$

or a bucket-order process $h \in \mathcal{K}^{(t_1, t_2)}$ with

$$\mathcal{K}^{(t_1, t_2)} = \mathcal{K}_{\mathcal{M}_{t_1}} \times \mathcal{K}_{\mathcal{M}_{t_1+1}} \times ... \times \mathcal{K}_{\mathcal{M}_{t_2}},$$

then likelihood evaluation would be fast. However, this is not well-evidenced in our setting: the consensus order from 1136 in Fig 13 contains the poset $H' = \{\langle 54, 57 \rangle, \langle 47, 57 \rangle, \langle 47, 58 \rangle\}$ in Fig 1, with each included edge having posterior probability 0.8 or above and absent edges below 0.5, so the true poset is probably not a VSP as it contains a suborder ismorphic to F.

7.3. *Test results.*   Bayes Factors $B_{\mathcal{V},\mathcal{H}}$ and $B_{\mathcal{K},\mathcal{H}}$ measuring the evidence for VSP poset-models and Bucket-Order models are defined in Sec. I.2. We estimate these on the same twelve short (five-year) time intervals we used for the Plackett-Luce analysis in Supp J.2 for accurate estimation. The seniority covariate is a near-constant in these short time intervals so seniority effects were set to $\beta_r = 0$, $r = 1, ..., S$ in the fitted poset model. The likelihood $p_{(U)}$ in Eqn 6 was used. For analysis in interval $[t_1, t_2]$, we take K, the dimension of $U_j^{(t)}$, to be $\lfloor \frac{1}{2} \max_{t \in [t_1, t_2]} m_t \rfloor$. The prior for ρ is Beta$(1, 1/6)$ and otherwise as Sec. 3.3.
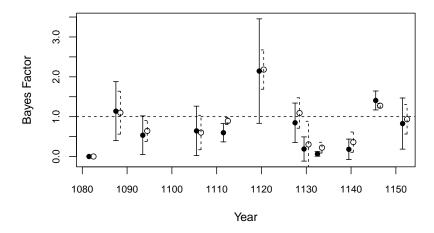

FIG 15. *Bayes Factors for VSP orders (dashed lines) and Bucket Orders (solid) fitted over the year intervals in Table 2. A value less than one is evidence against VSP or Bucket Orders. The x-axis value for each bar is the centre of the corresponding interval, $(t_1 + t_2)/2$. Error bars are two standard deviations.*

Results are given in Table 2 in Supp I and plotted in Fig 15. Each pair of points is an independent MCMC run. The prior probabilities $\pi(h \in \mathcal{V}^{(t_1,t_2)})$ and $\pi(h \in \mathcal{B}^{(t_1,t_2)})$ in Table 2 are surprisingly large given the sparsity of VSPs and Bucket Orders, so our prior for posets must favor VSPs and Bucket Orders. The rules forming VSP-orders and Bucket Orders compare groups rather than actors in a socially plausible way, so this may be a good thing. As Fig 15 shows, the data favour partial orders or are neutral, except for 1118-1122, so Bucket Orders and VSPs may be acceptable over some short time intervals. In some intervals VSPs and Bucket Orders are rejected (1080-1084 and around 1132-1134). The evidence for posets visible in Fig 15 will accumulate over longer time series.

8. Conclusions.   A new class of poset-models for time-series rank-data is summarised in Sec. 3.3. The latent variable poset-parameterisation in Eqn 15 made it straightforward to introduce a parameter controlling poset depth and incorporate actor-covariates informing the position of actors in the hierarchy. We fit the model to witness-list data in which the actors are eleventh and twelfth century bishops. In Sec. 6.2 we saw that the model recovered structure in the data which was anticipated by historians. In particular, the dependence of the status of a bishop on their seniority is clear in Fig 11. We checked for evidence that the depth parameter ρ, correlation θ and error probability p varied over time by looking at short time intervals (see Supp H) and found no evidence against our assumption of constant values over time. Further support comes from model comparison against a Placket-Luce mixture in Supp J.2 which favoured our model. The time-series extension of the Plackett-Luce model in Supp J.1 gave similar results for seniority effects and evolving authority, showing that the data overwhelm these model variations and conclusions are robust. The times-series Plackett-Luce model is fairly time consuming to fit so there was no great gain in efficiency over posets on our data.

We gave our main analysis in Sec. 6.3. This is the first quantitative analysis of these data and gave insights which historians find interesting. With few exceptions, witness-lists reflect precedence by diocese and seniority more than changing royal favour. We separated the effects of authority and seniority on status and confirmed (Johnson, 2013) that the bishops of London, Winchester and Lincoln had high authority and that Rochester had no special status. Personal authority changes in a few cases: the high status originally given to Nigel of Ely unwound, and Roger of Salisbury bucked the trend. This was known to historians. The apparent decline in authority of Norman bishops (as expressed in the lists) was unknown. Consensus partial orders (see Fig 13) gave an interpretable visualisation of the underlying social hierarchy in each year. The problem treated here appears in other guises: there is an extensive literature estimating animal dominance hierarchies with many similarities. For example, Foerster et al. (2016) studies a time-series of pairwise chimpanzee interactions and finds a role for seniority and evolution of status.

There is work to be done in computation and methodology to make these models more broadly applicable. First, our MCMC is time consuming (the experiments in this paper add up to about two years of MCMC if run as a single serial process). Whilst a careful analysis minimising any approximation of the target distribution is justified (there will never be any more for this period), scalable methods would be welcome, and VSP-orders may be an acceptable compromise from a modelling perspective. This may allow us to fit more complex noise models and treat "top-k" preference orders. Analysis of lists including lay witnesses (Jiang et al., 2023) required scalable methods in order to count linear extensions in partial orders over hundreds of elements. Application of our methods to general preference orders on thousands of items and thousands of lists is not presently feasible.

Second, many applications of poset-based models will require substantial model-building, paralleling the evolution of Mallows and Plackett-Luce models and including hierarchical models and mixture models for structured populations and clustering. The clerks who made the lists may have differed on status assessment so unrepresented group structure may be present. Some developments are given in Jiang and Nicholls (2024) (hereafter JN24). We worked with strong partial orders. Models for weak partial orders with ties are given in JN24. Also, our statistical model for "noise" in realised lists assumes errors occur in one direction only (queue jumping up or down, but not both). Noise models with bi-directional errors are explored in Jiang et al. (2023) and JN24. Statistical tools for selecting the number of features K in the status vector of a bishop would remove the need for robustness checks (our Supp G.3). This is addressed for fixed-time data in JN24 using reversible-jump MCMC.

Third, a small number of other covariates beside seniority are available in the data and might be explored in model elaboration. Covariates might enter the noise model also, to inform the probability and magnitude of displacements. One unexplored weakness of our error model is apparent in lists of length two: if the two bishops are ordered in a deep poset then the noise model assigns the same probability for the "wrong" order whether the two bishops are at each end of the poset (very different status) or adjacent in its transitive reduction (near-equal status). Displacements are probably "over-dispersed" in the noise-model.

From a historical perspective this study raises several questions. Complex precedence structures seem to have existed, but how were they known? Was there some kind of precedence handbook, or other means of transmission? Comparisons between patterns of precedence relations in pre-conquest lists and lists from later periods might be revealing. Later documents are more accurately dated and so a more fine-grained analysis may be possible. Finally, there are forgeries among acta of this period. Cases that go against the usual pattern may be an indicator of forgery.

Software used to carry out the analysis presented in this paper are available at

https://github.com/gknicholls/Partial-order-HMM-public.git.

REFERENCES

Arcagni, A., A. Avellone, and M. Fattore (2022). Complexity reduction and approximation of multidomain systems of partially ordered data. *Computational Statistics & Data Analysis 173*, 107520.

Asfaw, D., V. Vitelli, Ø. Sørensen, E. Arjas, and A. Frigessi (2017). Time-varying rankings with the Bayesian Mallows model. *Stat 6*(1), 14–30.

Bates, D. (1997). *Family Trees and the Roots of Politics: The Prosopography of Britain and France from the Tenth to the Twelfth Century*, Chapter The Prosopographical Study of Anglo-Norman Royal Charters, pp. 89–102. Boydell Press.

Bates, D. (1998). *The Acta of William I, 1066–1087*. Regesta regum Anglo-Normannorum 1066-1154. Oxford.

Bates, D. (2013). *The Normans and Empire*. Ford lectures. OUP Oxford.

Baxter, S. (2007). *The Earls of Mercia: Lordship and Power in Late Anglo-Saxon England*. Oxford University Press.

Beerenwinkel, N., N. Eriksson, and B. Sturmfels (2007). Conjunctive Bayesian networks. *Bernoulli 13*(4), 893–909.

Bengio, Y. and Y. Grandvalet (2004). No unbiased estimator of the variance of k-fold cross-validation. *Journal of Machine Learning Research 5*, 1089–1105.

Bogart, K. P. (1973a, 5). Maximal dimensional partially ordered sets I. Hiraguchi's theorem. *Discrete Mathematics 5*, 21–31.

Bogart, K. P. (1973b). Preference structures i: Distances between transitive preference relations. *The Journal of Mathematical Sociology 3*(1), 49–67.

Brightwell, G. (1993). *Surveys in Combinatorics*, Volume 187 of *London Mathematical Society Lecture Note Series*, Chapter Models of random partial orders, pp. 53–83. Cambridge Univeristy Press.

Brightwell, G. and P. Winkler (1991). Counting linear extensions. *Order 8*(3), 225–242.

Caron, F. and A. Doucet (2012). Efficient Bayesian inference for generalized Bradley-Terry models. *J. Comput. Graph. Statist. 21*(1), 174–196.

Caron, F. and Y. Teh (2012). Bayesian nonparametric models for ranked data. In F. Pereira, C. Burges, L. Bottou, and K. Weinberger (Eds.), *Advances in Neural Information Processing Systems*, Volume 25. Curran Associates, Inc.

Caron, F., Y. W. Teh, and T. B. Murphy (2014). Bayesian nonparametric Plackett–Luce models for the analysis of preferences for college degree programmes. *The Annals of Applied Statistics*, 1145–1181.

Clover, H. and M. Gibson (1979). *The Letters of Lanfranc, Archbishop of Canterbury*. Oxford.

Cronne, H. A. and R. H. C. Davis (1968). *Regesta Regis Stephani ac Mathildis Imperatricis ac Gaufridi et Henrici Ducum Normannorum, 1135–1154*. Regesta regum Anglo-Normannorum 1066-1154. Oxford.

Davis, H. and R. Whitwell (1913). *Regesta regum Anglo-Normannorum 1066-1154: Regesta Willelmi Conquestoris et Willelmi Rufi 1066-1100*. Regesta regum Anglo-Normannorum 1066-1154. Oxford.

Deng, K., S. Han, K. J. Li, and J. S. Liu (2014). Bayesian aggregation of order-based rank data. *Journal of the American Statistical Association 109*(507), 1023–1039.

Diaconis, P. (1988). Group representations in probability and statistics. *Lecture Notes-Monograph Series 11*, i–192.

Durante, D., D. B. Dunson, and J. T. Vogelstein (2017a). Nonparametric Bayes modeling of populations of networks. *Journal of the American Statistical Association 112*(520), 1516–1530.

Durante, D., D. B. Dunson, and J. T. Vogelstein (2017b). Rejoinder: Nonparametric Bayes modeling of populations of networks. *Journal of the American Statistical Association 112*(520), 1547–1552.

Dushnik, B. and E. W. Miller (1941). Partially ordered sets. *American Journal of Mathematics 63*(3), 600–610.

FEA (2022). Fasti Ecclesiae Anglicanae, 1066-1300. London: Institute of Historical Research; British History Online; Last accessed 22 October 2022; http://www.british-history.ac.uk/fasti-ecclesiae/.

Fishburn, P. C. and W. V. Gehrlein (1975). A comparative analysis of methods for constructing weak orders from partial orders. *J. Math. Sociology 4*(1), 93–102.

Fligner, M. A. and J. S. Verducci (1986). Distance based ranking models. *Journal of the Royal Statistical Society. Series B (Methodological) 48*(3), 359–369.

Foerster, S., M. Franz, C. M. Murray, I. C. Gilby, J. T. Feldblum, K. K. Walker, and A. E. Pusey (2016). Chimpanzee females queue but males compete for social status. *Scientific Reports 6*(1), 35404–35415.

Friedell, M. F. (1967). Organizations as semilattices. *American Sociological Review 32*(1), 46–54.

Froehlich, H., M. Fellmann, H. Sueltmann, A. Poustka, and T. Beissbarth (2007). Large scale statistical inference of signaling pathways from RNAi and microarray data. *BMC Bioinformatics 8*(386).

Gazes, R. P., R. R. Hampton, and S. F. Lourenco (2017). Transitive inference of social dominance by human infants. *Developmental Science 20*(2), e12367.

Gionis, A., H. Mannila, K. Puolamäki, and A. Ukkonen (2006). Algorithms for discovering bucket orders from data. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 561–566.

Given-Wilson, C. (1991). Royal charter witness lists 1327-1399. *Medieval Prosopography 12*(2), 35–93.

Glickman, M. E. and J. Hennessy (2015). A stochastic rank ordered logit model for rating multi-competitor games and sports. *Journal of Quantitative Analysis in Sports 11*(3), 131–144.

Guiver, J. and E. Snelson (2009). Bayesian inference for Plackett-Luce ranking models. In *proceedings of the 26th annual international conference on machine learning*, pp. 377–384.

Gwee, X. Y., I. C. Gormley, and M. Fop (2023). A latent shrinkage position model for binary and count network data. *Bayesian Analysis*, 1 – 29.

Haskins, G. L. (1938). Charter witness lists in the reign of king john. *Speculum 13*(3), 319–325.

Hiraguchi, T. (1951). On the dimension of partially ordered sets. *The science reports of the Kanazawa University 1*, 77–94.

Hunter, D. R. (2004). MM algorithms for generalized Bradley-Terry models. *The Annals of Statistics 32*(1), 384 – 406.

Irurozki, E., B. Calvo, and J. A. Lozano (2016). PerMallows: An R Package for Mallows and generalized Mallows models. *Journal of Statistical Software 71*(12), 1–30.

Irurozki, E., B. Calvo, and J. A. Lozano (2019). Mallows and generalized Mallows model for matchings. *Bernoulli 25*(2), 1160 – 1188.

Jiang, C. and G. K. Nicholls (2024). Non-parametric Bayesian inference for partial orders with ties from rank data observed with Mallows noise. https://arxiv.org/abs/2408.14661.

Jiang, C., G. K. Nicholls, and J. E. Lee (2023, 31 Jul–04 Aug). Bayesian inference for vertex-series-parallel partial orders. In R. J. Evans and I. Shpitser (Eds.), *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*, Volume 216 of *Proceedings of Machine Learning Research*, pp. 995–1004. PMLR.

Johnson, C. and H. A. Cronne (1966). *Regesta Henrici Primi : 1100-1135*. Regesta regum Anglo-Normannorum 1066-1154. Oxford.

Johnson, D. (2013, 10). Bishops and deans: London and the province of Canterbury in the twelfth century*. *Historical Research 86*(234), 551–578.

Kangas, K., T. Hankala, T. Niinimäki, and M. Koivisto (2016). Counting linear extensions of sparse posets. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, IJCAI'16, pp. 603–609. AAAI Press.

Kangas, K., M. Koivisto, and S. Salonen (2019). A faster tree-decomposition based algorithm for counting linear extensions. *Algorithmica*, 1–18.

Karzanov, A. and L. Khachiyan (1991). On the conductance of order Markov chains. *Order 8*, 7–15.

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association 90*(430), 773–795.

Keynes, S. (1980). *The Diplomas of King Aethlred "the Unready" 978–1016*. Cambridge Studies in Medieval Life and Thought: Third Series. Cambridge University Press.

Kleitman, D. J. and B. L. Rothschild (1975). Asymptotic enumeration of partial orders on a finite set. *Transactions of the American Mathematical Society 205*, 205–220.

Knuth, D. E. and J. L. Szwarcfiter (1974). A structured program to generate all topological sorting arrangements. *Information Processing Letters 2*(6), 153–157.

Licence, T. (2020, 09). *Edward the Confessor: Last of the Royal Blood*. Yale University Press.

Lu, T. and C. Boutilier (2014). Effective sampling and learning for Mallows models with pairwise-preference data. *J. Mach. Learn. Res. 15*(1), 3783–3829.

Luce, R. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology 15*(3), 215–233.

Luce, R. D. (1959). On the possible psychophysical laws. *Psychological review 66*(2), 81.

Mallows, C. L. (1957). Non-null ranking models. i. *Biometrika 44*(1/2), 114–130.

Mannila, H. (2008). Finding total and partial orders from data for seriation. In J.-F. Boucault (Ed.), *Discovery Science*, Volume 5255 of *LNAI*, Berlin Heidelberg, pp. 16–25. Springer-Verlag.

Mannila, H. and C. Meek (2000). Global partial orders from sequential data. In *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '00, New York, NY, USA, pp. 161–168. Association for Computing Machinery.

Martin, J. L. (2002). Some algebraic structures for diffusion in social networks. *The Journal of Mathematical Sociology 26*(3), 123–146.

McKeough, K. and M. Glickman (2024). Plackett–Luce modeling with trajectory models for measuring athlete strength. *Journal of Quantitative Analysis in Sports 20*(1), 21–35.

Meilă, M. and H. Chen (2016). Bayesian non-parametric clustering of ranking data. *IEEE Transactions on Pattern Analysis and Machine Intelligence 38*(11), 2156–2169.

Meilă, M. and H. Chen (2010). Dirichlet process mixtures of generalized Mallows models. In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, UAI'10, Arlington, Virginia, USA, pp. 358–367. AUAI Press.

Mogapi, O. (2009). *A Latent Partial Order Model for Social Networks*. Ph. D. thesis, University of Oxford.

Mollica, C. and L. Tardella (2017). Bayesian Plackett-Luce mixture models for partially ranked data. *Psychometrika 82*(2), 442–458.

Mollica, C. and L. Tardella (2020). PLMIX: an R package for modelling and clustering partially ranked data. *Journal of Statistical Computation and Simulation 90*(5), 925–959.

Muir Watt, A. (2015). *Inference for partial orders from random linear extensions*. Ph. D. thesis, University of Oxford.

Nicholls, G. K. and A. Muir Watt (2011). Partial order models for episcopal social status in 12th century England. *IWSM 2011*, 437.

Niinimäki, T., P. Parviainen, and M. Koivisto (2016). Structure discovery in Bayesian networks by sampling partial orders. *Journal of Machine Learning Research 17*(57), 1–47.

OEIS Foundation Inc, . (2022). The on-line encyclopedia of integer sequences. Published electronically at https://oeis.org.

Plackett, R. L. (1975). The analysis of permutations. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 24*(2), 193–202.

Rising, J. (2021). Uncertainty in ranking. https://arxiv.org/abs/2107.03459.

Roberts, J. M. (1990). Modeling hierarchy: Transitivity and the linear ordering problem. *The Journal of Mathematical Sociology 16*(1), 77–87.

Rousseau, J. and K. Mengersen (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73*(5), 689–710.

Russell, J. C. (1937). Social status at the court of King John. *Speculum 12*(3), 319–329.

Sakoparnig, T. and N. Beerenwinkel (2012). Efficient sampling for Bayesian inference of conjunctive Bayesian networks. *Bioinformatics 28*(18), 2318–2324.

Seshadri, A., S. Ragain, and J. Ugander (2020). Learning rich rankings. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin (Eds.), *Advances in Neural Information Processing Systems*, Volume 33, pp. 9435–9446. Curran Associates, Inc.

Sharpe, R., D. Carpenter, H. Doherty, M. Hagger, and N. Karn (2014). The charters of William II and Henry I. Online: Accessed October 2022; `actswilliam2henry1.wordpress.com/the-charters/`.

Shizuka, D. and D. B. McDonald (2012). A social network perspective on measurements of dominance hierarchies. *Animal Behaviour 83*(4), 925–934.

Sivula, T., M. Magnusson, and A. Nehtari (2022). Unbiased estimator for the variance of the leave-one-out cross-validation estimator for a Bayesian normal model with fixed variance. *Communications in Statistics - Theory and Methods*, 1–23.

Sørensen, Ø., M. Crispino, Q. Liu, and V. Vitelli (2020). BayesMallows: An R Package for the Bayesian Mallows Model. *The R Journal 12*(1), 324–342.

Tkachenko, M. and H. W. Lauw (2016). Plackett-Luce regression mixture model for heterogeneous rankings. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pp. 237–246.

Valdes, J. (1978). *Parsing Flowcharts and Series-Parallel Graphs*. Ph. D. thesis, Stanford, CA, USA. AAI7905944.

Valdes, J., R. E. Tarjan, and E. L. Lawler (1982). The recognition of series parallel digraphs. *SIAM Journal on Computing 11*(2), 298–313.

van Wietmarschen, H. (2022). What is social hierarchy? *Noûs 56*(4), 920–939. (page 926, *"In sum,..."*).

Vasconcelos, M. (2008). Transitive inference in non-human animals: An empirical and theoretical analysis. *Behavioural Processes 78*(3), 313–334.

Vehtari, A., A. Gelman, and J. Gabry (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing 27*(5), 1413–1432.

Vitelli, V., Øystein Sørensen, M. Crispino, A. Frigessi, and E. Arjas (2018). Probabilistic preference learning with the Mallows rank model. *Journal of Machine Learning Research 18*(158), 1–49.

Wells, M. B. (1971). Elements of combinatorial computing.

Winkler, P. (1985). Random orders. *Order 1*(4), 317–331.

BAYESIAN INFERENCE FOR PARTIAL ORDERS FROM RANDOM LINEAR EXTENSIONS: POWER RELATIONS FROM 12TH CENTURY ROYAL ACTA (SUPPLEMENTARY MATERIAL)

BY GEOFF K. NICHOLLS[7,g] JEONG EUN LEE[8,h] NICHOLAS KARN[9,i] DAVID JOHNSON [10,j] RUKUANG HUANG [11,k] AND ALEXIS MUIR-WATT [12,l]

[7]*Department of Statistics, The University of Oxford, UK.,* [g]*nicholls@stats.ox.ac.uk*

[8]*Department of Statistics, The University of Auckland, New Zealand ,* [h]*kate.lee@auckland.ac.nz*

[9]*Faculty of Arts and Humanities, University of Southampton, UK,* [i]*N.E.Karn@soton.ac.uk*

[10]*St Peter's college, The University of Oxford, UK ,* [j]*david.johnson@spc.ox.ac.uk*

[11]*Department of Psychiatry, University of Oxford, UK ,* [k]*rukuang.huang@jesus.ox.ac.uk*

[12]*Private researcher, London, UK ,* [l]*alexis.muirwatt@gmail.com*

APPENDIX A: DATA REGISTRATION AND WORKS BY HISTORIANS

A.1. Data sources and data-collection. The process behind the making of this dataset is complex. The acta were written by royal scribes, usually clergymen with legal training and political skill. Most were written at the behest of individuals or institutions who wanted to obtain rights or privileges or ensure the intervention of royal agents in their favour. The scribes handed over the documents to these individuals or institutions, and they were responsible for preserving them. There was no central governmental archive until the 1190s, so there were no copies kept at Westminster or elsewhere.

Over time, many acta were destroyed accidentally or purposefully; those which survive tend to come from institutions such as towns or great churches which had a stable existence over centuries, and which had the capacity to protect and manage an archive. The activities of manuscript collectors has also had a significant impact on what has survived. The acta are now mostly in the possession of the major research libraries and archives, with a small number in private hands.

Over the last century and a half, researchers have worked to identify where these documents survive, and to obtain transcripts of them. Editions and catalogues have been made from those collections (Davis and Whitwell, 1913; Johnson and Cronne, 1966; Cronne and Davis, 1968; Bates, 1998), and, more recently, databases (Sharpe et al., 2014). These are widely used by researchers and scholars and are key resources for major questions in political, social and cultural history.

A.2. Data registration. Our registration aimed to ensure that the data conformed to the observation process we describe in the paper. We clip the uncertainty ranges of the lists τ_i^-, τ_i^+, $i \in I$ so that they lie within the range allowed by the bishops they list, that is, we impose

$$[\tau_i^-, \tau_i^+] \subseteq \bigcap_{j=1}^{n_i} [b_{y_j}, e_{y_j}].$$

This can lead to lists being dropped. For example, the list with data-set id 677 (see below), which has a data-set year indicator of 1139 (with no uncertainty), contains Philip, Bishop of Bayeux. He was in post 1142-1163 so the uncertainty range would be clipped to the empty set. The list with data-set id 2627 has both Hugh and John, bishops of Lisieux, with an obvious

second list appended, starting with Henry I. These issues could be easily fixed. However, we simply dropped lists of this sort as the number was small (five).

The list with data-set id 2364 is a fairly typical example of a witness list we retained. The bishops do not quite appear as a group, as Ranulf is ahead of Maurice. We extract the sub-list of bishops in positions $3, 4, 5, 6, 7, 9$ and code them as numbers using the table in Fig 24 yielding the registered list $Y = (23, 19, 30, 32, 5, 18)$.

List id 2364 Year range [1107,1108], Witness List: [1] Matilda II, wife of Henry I, queen of England; [2] Anselm, abbot of Bec, archbishop of Canterbury; [3] Robert, Bloet, bishop of Lincoln; [4] Robert, de Limesey, bishop of Chester; [5] John, Bishop of Lisieux; [6] Richard, Bishop of Bayeux; [7] Gundulf, bishop of Rochester; [8] Ranulf, chancellor of Henry I; [9] Maurice, bishop of London; [10] William, de Roumare, earl of Lincoln 1140; [11] Robert, II, count of Meulan; [12] David, King of Scots; [13] Robert, de Ferrers, earl of Derby.

List id 2627 Year range [1130,1135], List: [1] Odo, Stigandus; [2] Osmund, Boenot; [3] Serlo, de Mansione, Malgerii; [4] William, de Mirebel; [5] Hugh, Buscard; [6] Ranulf, de Iz; [7] Grento, de Vals; [8] Ralph, de Vals; [9] William I, king of England; [10] John, archbishop of Rouen; [11] Hugh, bishop of Lisieux; [12] Michael, bishop of Avranches; [13] Durand, abbot of Troarn; [14] Ainard, abbot of St Mary's, Dives; [15] Nicholas, abbot of St Ouen; [16] Roger, de Montgomery, earl of Shrewsbury; [17] Roger, de Beaumont; [18] William, de Breteuil, count; [19] Henry I; [20] Matilda II, wife of Henry I, queen of England; [21] John, Bishop of Lisieux; [22] Rabel, de Tancarville, chamberlain; [23] Thurstan, Archbishop of York; [24] Robert, earl of Gloucester; [25] William, de Warenne II, earl of Surrey d. 1138; [26] Robert, de Beaumont, earl of Leicester; [27] Payn, Peverel.

List id 677 Year range [1139,1139], Witness List: [1] Robert, Losinga, bishop of Hereford; [2] Philip, Bishop of Bayeux; [3] Roger, archdeacon of Fecamp, temp. Stephen; [4] Walter, archdeacon of Oxford; [5] Waleran, count of Meulan; [6] Ingelram de Say, temp. Stephen; [7] Walter, of Salisbury, temp. Stephen; [8] Robert, de Vere; [9] William, de Pont de, l'Arche.

A.3. Dioceses of interest. The data display bishops with 31 distinct diocese names: Lincoln, Durham, Chester, Sherborne, Winchester, Chichester, Bayeux, Lisieux, Evreux, Sees, Avranches, Coutances, Exeter, London, Rochester, Worcester, Salisbury, Bath, Thetford, Wells, Le Mans, Hereford, Bangor, Ely, St Davids, Norwich, Carlisle, St Asaph, Tusculum, the Orkneys and Llandaff. We excluded the Bishop of Tusculum who appeared at the end of two long bishop-lists on the grounds that they they were not part of the group of interest (they would have been listed according to other criteria and in fact seem to have been regarded that way by the clerks as they were placed outside the group of bishops). We removed bishops who only appeared in a single list. Depending on the period $[t_1, t_2]$ of interest, this could reduce an included list from length two to one and removing that list could reduce the number of lists a bishop appeared in. We repeated this thinning to convergence. For the period $[1080, 1155]$ this removed 3 lists of lengths 2, 2 and 3 so the number of lists went down from 374 to 371. In addition, six lists lost one bishop and one lost 3. It reduced the number of bishops from 81 to 67. The bishops of six dioceses, Sherborne, Wells, Le Mans, St Asaph, Tusculum and The Orkneys witnessed no lists in the final data set.

A.4. Discussion of data registration. Removing bishops from lists in this way was done to avoid carrying out inference for a subset of latent U-process parameters (see Sec. 3.2) which the data don't inform as we would be integrating out near-prior marginals using MCMC. If we include them, the U-dimension is $\dim(U) = K \times 1408$ (with $K = 11$ in Sec. 6.3 and $K = 22$ in Supp G.3). Removing them reduces the latent dimension to

$\dim(U) = K \times 1223$, a 13% reduction in the number of continuous variables. The prior is marginally consistent, so in this respect the analysis is unchanged. We loose 3 of 374 lists and 7 lists are shortened, four loosing their last element.

Removing bishops from longer lists could remove information informing the parameters of the bishops remaining (in all the models we considered, including the poset and Plackett-Luce models). If we start with a list $Y = (j_1, ..., j_n)$ with n large then this supports the edge $\langle j_1, j_n \rangle$ relatively strongly. If the list is reduced to a list (j_1, j_n) then the edge $\langle j_1, j_n \rangle$ is more weakly supported. If we have two lists (j_1, j_2) and (j_2, j_3) then we have support for $\langle j_1, j_3 \rangle$. If we remove bishop j_2 then we loose this information. These examples suggest that dropping bishops makes our estimated posets less certain, and contain fewer order relations. However, dropping bishops can remove evidence against orders. If we have three lists (j_1, j_2), (j_2, j_3) and (j_3, j_1) then there is evidence for no order between j_1 and j_3. If we remove j_2 then we are left with evidence for the order $\langle j_3, j_1 \rangle$.

In our setting 98% of the lists were unchanged so the effect of thinning will be slight. In an earlier version of this paper we made a more substantial thinning of the data, dropping eleven dioceses. This reduced the the number of bishops to 59 and shortened 10% of the lists. Results were very similar to those here, down to small details (such as β_{11} being "out of order" in Fig 11 and the authority curves in Fig 14 for matching dioceses being indistinguishable). Over the 76 years 1080-1155 we estimate around 13,000 linearly independent posterior probabilities for relations between bishops present in both analyses. The median absolute difference in estimated posterior probabilities $|\hat{\xi}^{(1,t)}_{\langle j_1, j_2 \rangle} - \hat{\xi}^{(2,t)}_{\langle j_1, j_2 \rangle}|$ in Eqn 25 for relations between bishops in the old and new analyses ("1" and "2") was approximately 3%.

A.5. Works by historians analysing witness lists. Historians have used the evidence of witness lists in a number of ways: (a) to determine the identity of those who were close to the grantors of documents (on the assumption that witnesses were likely present with the person in whose name a document was issued); (b) to trace the career of individual witnesses (since if a witness appears after a certain date with a certain title, but appears without that title previously, it is reasonable to assume that that title was acquired before the date of the document in which it is first mentioned, and after those earlier ones in which it is not); and (c) as the basis for a statistical analysis of the witnesses mentioned in the lists. Statistical techniques have, however, scarcely been used at all.

The first modern statistical analysis of royal witness lists was published by Russell (1937). Russell's aims resembled those of the current research since he attempted to use the evidence of witness lists from royal charters issued by King John (king of England 1199 – 1216) to determine the relative status of individuals mentioned in those lists. This started a debate. In 1938, a further article was published (Haskins, 1938) which questioned Russell's statistical methodology and thus, in turn, suggested his conclusions were unsound. Haskins writes "the editor of the charter rolls for the reign of John, saw fit to deny explicitly the suggestion that there was any order among the witnesses which might have been dictated by considerations of rank or precedence" and "the precedence of witnesses to private grants is too erratic to serve as an index to their respective station". We revisited Russell (1937) and find the statistical reasoning (a permutation test) is incomplete rather than incorrect (a statistic measuring the dispersion of random permutations is compared to its value on data; they differ as anticipated but the variation isn't quantified). However, the point being made in the single paragraph which contains numbers, that the lists are too similar to be generated uniformly at random, is so obvious it is barely worth justifying. We agree that the lists are a noisy indicator of precedence. However, Haskins did not have available the tools of modern statistical inference for quantifying uncertainty in the structure behind the noise.
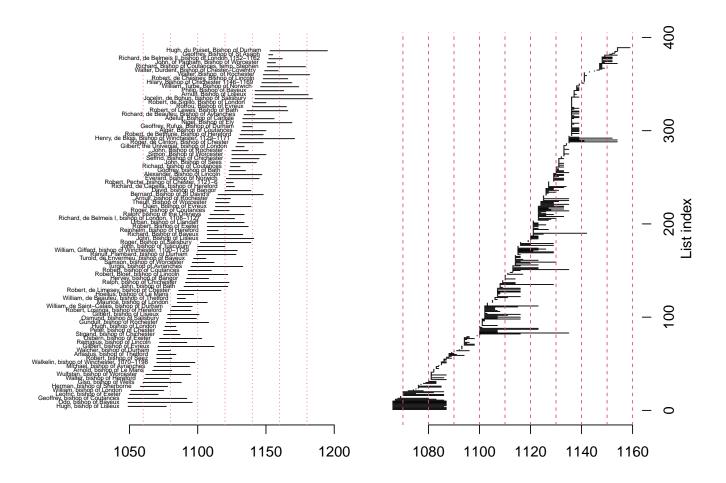
FIG 16. *Bishop (left) and list (right) date intervals. Each horizontal bar above left spans the interval $[b_j, e_j]$ over which the indicated bishop $j \in \mathcal{M}$ held their appointment. Each bar above right spans the uncertainty interval $[\tau_i^-, \tau_i^+]$ for list $i \in \mathcal{I}$.*

Witness lists did continue to be the subject of analysis, although this has been heavily biased towards attempts to discover the names of those most prominent in the royal household or administration. A seminal article (Given-Wilson, 1991) on royal acta from the reigns of King Edward III (king of England 1327 –1377) and King Richard II (king of England 1377 – 1399) provided some exception to this. Given-Wilson provided a quantitative as well as qualitative analysis; but, although he noted the general fact that individuals were grouped together into groups related to their status and societal role (bishops, earls and so on), Given-Wilson's concern was not with relative status but rather with the frequency with which individuals attested royal charters. He thus produced a series of tables in which individuals were grouped by category and in which the frequency of their appearance was measured.

Historians of the period before the Norman Conquest in 1066, and of the years which immediately followed it, have paid closer attention to the ways in which witness lists can be used to illustrate questions of status. However, as with historians of the later period, they have not applied statistical techniques to the question. Amongst those who have considered the issue, Keynes (1980) noted the fact that whilst again documents from the reign of King Æthelred II (king of England 978 – 1016) demonstrate the fact that particular groups of individuals always appeared together, these groups did not always appear in the same order;

Baxter (2007) reviewed the relative positions amongst the earls in a similar period, and noted that the comparative positions of one earl in relation to another was subject to change as a reflection of political instability. In a discussion more directly relevant to the current article, Licence (2020) discussed the place of bishops in witness lists produced under King Edward the Confessor (king of England 1042 – 1066). Licence detected a role for bishops in the production of documents for beneficiaries in their diocese, and suggested that generally speaking their involvement influenced their place in the list with the local bishop 'concluding the senior group' of witnesses; where the diocesan was not responsible for the drafting, Licence suggested that the man who was might occupy this position.

In a review of royal acta produced following the Conquest, Bates (1997) stressed the impact of a document's diplomatic and whether its beneficiary was sited in Normandy or England. Again, Bates observed the probability that witnesses were not always physically present when certain types of document were drawn up (but, it is worth noting that this need not have influenced the order in which witnesses were inscribed). Bates further noted that a 'more probing analysis – impossible to undertake here – would examine closely the pattern of attestations'. It is the purpose of this paper, albeit for a slightly later period, to begin that analysis. In particular, to examine the ways in which the witness lists of royal acta reflect the status and position of one particular group – the bishops – in the period under discussion. At the Council of London of 1075, it was ruled that English bishops should enjoy the following order of precedence: Canterbury, York, London and then Winchester, followed by other bishops in the order of their appointment. We seek here to examine whether or not the witness lists reflect any such order. Later in the middle ages, the order of precedence was further extended to the bishops of other sees. We thus also seek to show whether or not at this earlier date there were any indications that such an extended order was already being considered.

APPENDIX B: PROPERTIES OF THE OBSERVATION MODEL

B.1. Context dependence. Our queue-based idealisation of the witness process in Sec. 3.1.2 led us to a context dependent model for random orders. For example, the poset $H = \{\langle 2,3 \rangle\}$ on $[M] = \{1,2,3\}$ has three linear extensions $\mathcal{L}[H] = \{(1,2,3),(2,1,3),(2,3,1)\}$ so if all the actors are present ($O = [M]$) then the probability actor 1 comes before 2 in a random list is one third. However, if only 1 and 2 are present ($O = \{1,2\}$) then $H[O] = \emptyset$ and $\mathcal{L}[\emptyset] = \{(1,2),(2,1)\}$ so 1 beats 2 with probability one half. This is called context dependence as the presence of actor 3 changes the probability for 1 come ahead of 2. Plackett-Luce models, which satisfy the Choice Axiom of Luce (Luce, 1977), are context independent.

B.2. Counting Linear extensions. In this Supp, which picks up from Sec. 3.1.2, we outline how we count linear extensions of a given poset $H \in \mathcal{H}_{[m]}$.

Recall from Sec. 3.1.2 that $C(H)$ is the number of linear extensions of poset H and $C_j(H)$ is the number that start with j. If $\mathcal{T}(H) = \{j \in [m] : C_j(H) > 0\}$ is the set of "top elements" of H then

$$C(H) = \sum_{j \in \mathcal{T}(H)} C_j(H).$$

This is computed using a suborder recursion. If $j \in \mathcal{T}(H)$ and we let

(23) $$H_{-j} = H[[m] \setminus \{j\}]$$

give the poset H with vertex j removed then $C_j(H) = C(H_{-j})$, so we can prune away the vertices one at a time, starting at the top of the poset. The count is available in closed form (or rapidly) for some subclasses of posets including the Bucket Orders and vertex-series-parallel orders (see Sec. 7). For example, the empty poset, with $H = \emptyset$, has $C(H) = m!$. We implemented a simple R program using a recursion which terminates at these "easy" cases. We have also faster code based on Kangas et al. (2019) which we used in later work. Links to these are shared in our online code.

B.3. Proof of Proposition 1. The "queue jumping" error model reduces to the uniform distribution on linear extensions of the poset when the probability to jump up the queue (the noise probability p) is equal zero.

Proposition 1. The noisy free and noisy models (Eqns 4, 5 and 6) all coincide at $p = 0$.

PROOF. Take $H \in \mathcal{H}_{[m]}$ and $Y \in \mathcal{L}[H]$ so that $C(H(Y_{j:m})) > 0$ for each $j = 1, \ldots, m$ (a poset has at least one linear extension by Dushnik and Miller (1941)). Substituting $p = 0$ in Eqn 5 and expanding,

$$p_{(D)}(Y|H, p = 0) = \frac{C_{Y_1}(H[Y_{1:m}])}{C(H[Y_{1:m}])} \times \frac{C_{Y_2}(H[Y_{2:m}])}{C(H[Y_{2:m}])} \times \ldots \times \frac{C_{Y_{m-1}}(H[Y_{m-1:m}])}{C(H[Y_{m-1:m}])}$$

$$= C(H[Y_{1:m}])^{-1}$$

which is $P(Y|H)$ in Eqn 4. The second line follows because the last numerator factor is equal one, $C_{Y_{m-1}}(H[Y_{m-1:m}]) = 1$ (the poset $H[Y_m]$ we get when we remove Y_{m-1} is a poset with just a single node) and $C_{Y_j}(H[Y_{j:m}]) = C(H[Y_{j+1:m}]) \neq 0$ (the number of linear extensions of $H[Y_{j:m}]$ that start with Y_j appearing in the numerator is equal to the number of linear extensions of $H[Y_{j+1:m}]$, the same poset but with Y_j removed, and this is in the denominator of the next factor) and so the product of counts is telescoping.

When $Y \notin \mathcal{L}[H]$ then there is $j \in [m]$ such that $C_{Y_j}(H[Y_{j:m}]) = 0$ (one of the actors was chosen out of order, so it didn't head any linear extension of $H[Y_{j:m}]$). In this case we must

have $C_{Y_j}(H[Y_{j:m}]) \neq C(H[Y_{j+1:m}])$ because $C(H[Y_{j+1:m}]) > 0$ always. Since at least one numerator factor is zero, and all the denominator factors are positive, $p_{(D)}(Y|H, p = 0) = 0$ in this case. It follows that $p_{(D)}(Y|H, p = 0) = p(Y|H)$ for all $Y \in \mathcal{P}_{[m]}$.

The proof that $p_{(U)}(Y|H, p = 0) = p(Y|H)$ is similar. □

APPENDIX C: PROPERTIES OF PRIORS

C.1. Proof of Proposition 2. When $\beta = 0_S$ we have $Z_j^{(t)} = U_j^{(t)} = Z(U_j^{(t)}, 0_S; s)$ for each $t \in [B, E]$ and each $j \in \mathcal{M}_t$. Let

$$\pi_{\mathcal{H}^{(B,E)}}(h|\rho, \theta, \beta = 0_S) = \int \mathbb{I}_{h=h(Z(U, 0_S; s))} \pi(U|\rho, \theta) \, dU$$

give the marginal distribution of h given ρ, θ and $\beta = 0_S$. The marginal distribution of h is

$$(24) \qquad \pi_{\mathcal{H}^{(B,E)}}(h|\beta = 0_S) = \int_{[0,1]^2} \pi_{\mathcal{H}^{(B,E)}}(h|\rho, \theta, \beta = 0_S) \pi(\rho, \theta) \, d\rho d\theta.$$

We can remove $j \in \mathcal{M}$ "from the beginning" or from the realised poset "at the end" and get the same random poset. In the former, the random poset at time t is $h(Z_{-j}^{(t)})$ where $Z_{-j}^{(t)}$ is an $m_t - 1 \times K$ matrix with one independent row $Z_{j'}^{(t)}$ for each $j' \in \mathcal{M}_t \setminus \{j\}$. Let $h(Z_{-j}) = (h(Z_{-j}^{(t)}))_{t=B}^E$ be a time series of partial orders realised in this way. In the latter, we take a suborder at the end. Let $\mathcal{M}_t^{-j} = \mathcal{M}_t \setminus \{j\}$ and let $h_{-j}^{(t)} = h(Z^{(t)})[\mathcal{M}_t^{-j}]$ be the suborder at time t when we remove j from the order realised on the full Z-matrix. Let $h_{-j} = (h_{-j}^{(t)})_{t=B}^E$ give the time series. Let

$$\mathcal{H}_{-j}^{(B,E)} = \mathcal{H}_{\mathcal{M}_B^{-j}} \times \mathcal{H}_{\mathcal{M}_{B+1}^{-j}} \times ... \times \mathcal{H}_{\mathcal{M}_E^{-j}},$$

so that $h(Z_{-j})$ and h_{-j} are both in $\mathcal{H}_{-j}^{(B,E)}$. We begin by stating the simple property really underpinning marginal consistency.

PROPOSITION 4. The random poset processes obtained by mapping the full Z-process to a poset process and removing $j \in \mathcal{M}$ from each poset in which it appears ("at the end"), or removing j from the Z-process ("from the beginning") are equal, so $h(Z_{-j}^{(t)}) = h(Z^{(t)})[\mathcal{M}_t^{-j}]$ for $t \in [B, E]$.

PROOF. This follows from the fact that the relations $\langle j_1, j_2 \rangle$ in h determined by Eqn 10 are not affected by the presence or absence of a third element j. □

Proposition 2. Let $g \in \mathcal{H}_{-j}^{(B,E)}$ be given. The prior in Eqn 24 is marginally consistent,

$$\pi_{\mathcal{H}_{-j}^{(B,E)}}(g|\beta = 0_S) = \sum_{h \in \mathcal{H}^{(B,E)}} \mathbb{I}_{g=h_{-j}} \pi_{\mathcal{H}^{(B,E)}}(h|\beta = 0_S).$$

for each $j \in \mathcal{M}$.

PROOF. Marginal consistency follows from Proposition 4 using a Chinese Restaurant Process style construction. Let $\mathcal{M} = \{j_1, ..., j_M\}$ (labelling in any order). Add each actors ("customer") $j_1, ..., j_M$ one at a time, using the fact that the Z_j-processes are jointly independent for $j \in \mathcal{M}$.

1. Initialise:

    a) Fix $K \geq 1$, simulate $(\rho, \theta) \sim \pi(\rho, \theta)$ and set $\beta = 0_S$.

    b) For $t \in [B, E]$ set $\mathcal{M}_{t,1} = \{j_1\} \cap \mathcal{M}_t$, $h_{(1)}^{(t)} = \emptyset$ and simulate $Z_{j_1} \sim \mathrm{VAR}_{K,\rho,\theta}^{(b_{j_1}, e_{j_1})}(1)$
    (recall $\beta = 0_S$).

2. For $i = 2, ..., M$ do

    a) Simulate $Z_{j_i} \sim \mathrm{VAR}_{K,\rho,\theta}^{(b_{j_i}, e_{j_i})}(1)$.

    b) For $t \in [B, E]$ set $\mathcal{M}_{t,i} = \{j_1, ..., j_i\} \cap \mathcal{M}_t$ and construct $h_{(i)}^{(t)}$ by adding to $h_{(i-1)}^{(t)}$
    the order relations between j_i and $j_1, ..., j_{i-1}$, as follows: For $j \in \{j_1, ..., j_{i-1}\}$ the
    order relation $\langle j_i, j \rangle$ (resp. $\langle j, j_i \rangle$) is added if $Z_{j_i,k}^{(t)} > Z_{j,k}^{(t)}$ (resp. $Z_{j,k}^{(t)} > Z_{j_i,k}^{(t)}$) for each
    $k = 1, ..., K$.

3. Set $h^{(t)} = h_{(M)}^{(t)}$ for each $t \in [B, E]$ and $h = (h^{(t)})_{t=B}^{E}$.

The final partial-order time series $h \sim \pi_{\mathcal{H}^{(B,E)}}(\cdot | \beta = 0_S)$ does not depend on the order in which actors $j_1, ..., j_M$ arrive so we may make $j_M = j$ the last arrival. Also, by Proposition 4, $g = h_{-j}^{(t)} = h_{(M-1)}^{(t)}$ is the poset state of the process for each $t \in [B, E]$ before $j_M = j$ arrives. By construction, $g \in \mathcal{H}_{-j}^{(B,E)}$ and $g \sim \pi_{\mathcal{H}_{-j}^{(B,E)}}(\cdot | \beta = 0_S)$. If we stop the process at step $i = M - 1$ then the poset we obtain must be the marginal over all continuations of the process to the next step, giving Proposition 2. $\qquad\square$

Every poset $h_{(i)}^{(t)}$, $i = 1, ..., M - 1$ is a suborder of the posets which come after it in the actor sequence, and adding or removing an actor does not change the order relations between other actors who have already arrived. This may be surprising, as we might expect new relations to add a cascade of relations by transitive closure. However, it is straightforward, as the relation between Z_{j_1} and Z_{j_2} in Eqn 10 isn't affected by the presence or absence of Z_{j_3}.

We removed covariates in the analysis above. In some settings, covariate levels $s_{t,j}$, $j \in \mathcal{M}_t$ characterise relations *between* actors active at time t and would change if we add or remove a actor. However, if covariate levels $s_{t,j}$ are intrinsic then they don't change when we add and remove actors. In that case $Z_{j_1}^{(t)} = U_{j_1}^{(t)} + 1_K \beta_{s_{t,j_1}}$ doesn't change when we add or remove actor j_2 and so relations between j_1 and a third actor j_3 don't change. It is then straightforward to show that $\pi_{\mathcal{H}^{(B,E)}}(h | \beta)$ is consistent for every β and so $\pi_{\mathcal{H}^{(B,E)}}(h)$ is also.

C.2. Proof of Proposition 3. In this section we prove the universal expression property claimed in Sec. 4.2.

Proposition 3. *Suppose $\min_{t \in [B,E]} m_t \geq 4$. The probability $\pi_{\mathcal{H}^{(B,E)}}(h)$ in Eqn 19, given by the generative model Eqn 15 with $K \geq \lfloor D/2 \rfloor$, assigns a positive probability mass $\pi_{\mathcal{H}^{(B,E)}}(h) > 0$ to every time-series $h \in \mathcal{H}^{(B,E)}$.*

Before giving a proof we note that the condition $\min_{t \in [B,E]} m_t \geq 4$ requires the number of active actors to always exceed four. This condition is needed for Hiraguchi's Theorem to hold. See Hiraguchi (1951) for the original result and Bogart (1973a) for an accessible proof and discussion. In our application in Sec. 2 $m_t \geq 11$ so the condition is satisfied.

PROOF. We first show that *(1) every poset in the time series h has dimension at most K.* A total order is a poset in which every pair of elements is ordered. The intersection of K total orders $\ell = (\ell^{(1)}, ..., \ell^{(K)})$ on the ground set $[m]$ is a poset $H = \cap_{k=1}^{K} \ell^{(k)}$ on m elements satisfying $H \in \mathcal{H}_{[m]}$. A given poset may be decomposed into total orders in many different ways (and always at least one, Dushnik and Miller (1941)). For example, the poset H in Fig 6 has three linear extensions, $Y^{(1)} = (1, 2, 3, 4, 5), Y^{(2)} = (1, 2, 4, 3, 5)$

and $Y^{(3)} = (1, 4, 2, 3, 5)$. We can turn these into total orders in the obvious way, setting $\ell^{(i)} = \{\langle Y_j^{(i)}, Y_{j'}^{(i)} \rangle \in Y^{(i)} \times Y^{(i)} : j < j'\}$, $i = 1, 2, 3$. Then H is both $H = \ell^{(1)} \cap \ell^{(3)}$ and $H = \ell^{(1)} \cap \ell^{(2)} \cap \ell^{(3)}$.

The dimension of a poset $\dim(H)$ is the smallest number of total orders that intersect to give H. It is known (Hiraguchi, 1951) that if $H \in \mathcal{H}_{[m]}$ then $\dim(H) \leq \lfloor m/2 \rfloor$. Restoring time to the notation, since $D = \max_t m_t$, we can take any $K \geq \lfloor D/2 \rfloor$ and be sure that $\dim(h^{(t)}) \leq K$ for every $t \in [B, E]$ and all $h \in \mathcal{H}^{(B,E)}$, where $\mathcal{H}^{(B,E)}$ (defined in Eqneq:h-sequence-space-defn) is the set of all possible poset time series $h = (h^{(B)}, h^{(B+1)}, \ldots, h^{(E)})$ on the sequence of ground sets $\mathcal{M}_B, \mathcal{M}_{B+1}, \ldots, \mathcal{M}_E$.

We next show that *(2) the mapping $h^{(t)} = h(Z^{(t)})$ takes the intersection over columns of the rank order for each column of $Z^{(t)}$.* Consider a $m_t \times K$ matrix $Z^{(t)}$ in a particular year $t \in [B, E]$. The rule mapping $Z^{(t)}$ to $h^{(t)}$ in Eqn 10 just takes the intersection of the rank orders of the columns of $Z^{(t)}$: for $k = 1, ..., K$, the rank order in column k is

$$\ell^{(t,k)} = \{\langle i, j \rangle \in \mathcal{M}_t \times \mathcal{M}_t : Z_{i,k}^{(t)} > Z_{j,k}^{(t)}\}$$

so Eqn 10 is the same as

$$h(Z^{(t)}) = \cap_{k=1}^K \ell^{(t,k)}.$$

Write $\ell^{(t)} = (\ell^{(t,k)})_{k=1}^K$ for the K total orders and write $\ell^{(t)} = \ell(Z^{(t)})$ to show that $\ell^{(t)}$ is determined from $Z^{(t)}$.

We next show *(3) for any partial-order time series h there is a set of latent-variable time series Z each realising h. The set has non-zero measure.* We realise any particular time series $h \in \mathcal{H}_{[A,B]}$ if we realise Z such that $h(Z) = h$, so we realise h if we realise a sequence of $Z^{(t)}$-matrices with column rank orders $\ell^{(t)} = \ell(Z^{(t)})$ that intersect to give $h^{(t)}$. Since $\dim(h^{(t)}) \leq K$ we know that such a sequence of rank orders exists, so we get h with positive prior probability if the probability to realise such a Z is positive.

The prior for Z is given in terms of a continuous mixture (over ρ and θ) of multivariate normal U and β prior densities. It has a density with respect to Lebesgue measure on the space

$$\mathcal{Z}^{(B,E)} = \mathbb{R}^{Km_B} \times \mathbb{R}^{Km_{B+1}} \times ... \times \mathbb{R}^{Km_E}$$

which is strictly positive for all $Z \in \mathcal{Z}^{(B,E)}$ (so every possible sequence of Z-matrices can be realised). Let $h \in \mathcal{H}^{(B,E)}$ be given and let $\ell^{(t)}$ be any decomposition of $h^{(t)}$ into K total orders (at least one decomposition exists, and if $\dim(h^{(t)}) < K$ then we just repeat total orders as this does not change their intersection). Let

$$\mathcal{Z}_{t,\ell} = \{z \in \mathbb{R}^{Km_t} : \ell(z) = \ell^{(t)}\}$$

be the set of Z-matrices in $\mathbb{R}^{Km_t}$ representing year t that yield the desired column rank orders.

The sets $\mathcal{Z}_{t,\ell}$ are not empty: we can take the m_t entries in the k'th column $Z_{\mathcal{M}_t,k}^{(t)} = (Z_{j,k}^{(t)})_{j \in \mathcal{M}_t}$ of $Z^{(t)}$ to be any set of real numbers that match the column rank order constraint imposed by $\ell^{(t,k)}$. It follows that the volume measure of $\mathcal{Z}_{t,\ell}$ is strictly greater than zero.

Any $z \in \mathcal{Z}_{t,\ell}$ satisfies $h(z) = h^{(t)}$. However $\ell^{(t)}$ above is just one decomposition of $h^{(t)}$ into K (not necessarily distinct) total orders, so $Z_{t,\ell}$ will typically be smaller than the set of all z's giving $h^{(t)}$. Let

$$\mathcal{Z}(h^{(t)}) = \{z \in \mathbb{R}^{m_t K} : h(z) = h^{(t)}\}$$

be that set, with $\mathcal{Z}(h^{(t)}) \supseteq \mathcal{Z}_{t,\ell}$. Let

$$\mathcal{Z}(h) = \mathcal{Z}(h^{(B)}) \times \mathcal{Z}(h^{(B+1)}) \times ... \times \mathcal{Z}(h^{(E)}),$$

and

$$\mathcal{Z}_\ell = \mathcal{Z}_{B,\ell} \times \mathcal{Z}_{B+1,\ell} \times ... \times \mathcal{Z}_{E,\ell}$$

so that again, $\mathcal{Z}(h) \supseteq \mathcal{Z}_\ell$ and the measure of $\mathcal{Z}_\ell$ is not zero.

Finally, *(4) the prior probability to realise any time series $h \in \mathcal{H}^{(B,E)}$ is not zero.* Indeed,

$$\pi_{\mathcal{H}^{(B,E)}}(h) = \Pr(Z \in \mathcal{Z}(h))$$
$$\geq \Pr(Z \in \mathcal{Z}_\ell)$$
$$> 0,$$

as the density of Z is not zero on the set $\mathcal{Z}_\ell$ and this set has non-zero measure. □

APPENDIX D: PRIOR AND POSTERIOR DEPTH DISTRIBUTIONS

In this section we report some summaries from prior simulation of depth distributions and compare them with posterior depth distributions. We have said that we would like prior depth distributions to be reasonably uniform. In Fig 17 we plot prior depth distributions. There is one curve for each of the $T = 76$ years showing the prior depth distribution for that year. The two plots show distributions for $K = 11$ and $K = 22$ (left and right respectively).

In Sec. 3.2.3 we mentioned that the uniform prior on posets $H \sim \text{Unif}(\mathcal{H}_{[m]})$ concentrates on posets of depth three as $m \to \infty$ (Kleitman and Rothschild, 1975). We illustrate that along with the depth distributions of our priors. In Fig 17 (left and right), the dashed curve shows the prior depth distribution determined by the uniform prior. A random poset $H \sim \text{Unif}(\mathcal{H}_{[m]})$ simulated using the MCMC algorithm given in Muir Watt (2015) is shown in Fig 20. This illustrates the rather dramtic weighting present in this prior.

Our own prior distributions tend to rule out partial orders of depth 1 and are otherwise reasonably flat, weighting somewhat in favour of lower depth. Taking $K = 22$ raises the prior probability for depth 1 slightly. This is hard to see in Fig 17 but visible in the left panels in Figs 18 and 19. Fig 18 shows prior and posterior depth distributions for the analysis in Sec. 6.3 with $K = 11$. We see that the whiskers do not extend to depth 1 in all years in the prior plot at left. Fig 19 shows prior and posterior depth distributions for the analysis in Supp G.3 which is the same as that in shown in Fig 18 except that $K = 22$. We see that the whiskers do now extend to depth 1 in all years in the prior plot at left.

The prior we have given has reasonably flat support over depths that the data actually supports. We may be concerned that the depths with highest posterior probability in Figs 18 and 19 coincide with relatively higher prior-probability depths in Fig 17 (so, around 2-5). However, this is not prior domination. The data easily dominates the prior. We show this in Supp K.2 where we take synthetic data with the same list content (so the same o_i, $i \in \mathcal{I}$ as the real data, but simulated data y_i', $i \in \mathcal{I}$). We take "true" partial orders which are complete (or near complete) orders so the true depth is around m_t, $t \in [B, E]$ for each year. We reconstruct this depth well, so the prior is in no way dominating the reconstructed depth. At the end of Supp K.2 we remark on the corresponding issue of potential prior bias at depth 1.
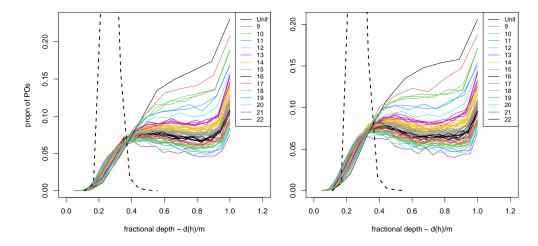

FIG 17. *Monte-Carlo estimates of the prior depth distribution for the poset at each time - using the prior for h summarised in Sec. 3.3: (Left) $K = 11$ columns in the Z-matrix; (Right) $K = 22$. This distribution varies as the number of bishops (indicated by color) in the poset varies over time. The x-axis shows relative depth $d(h^{(t)})/m_t$ (equal one for a total order). The dashed curve is the depth distribution for a uniform poset with $m_t = 18$.*
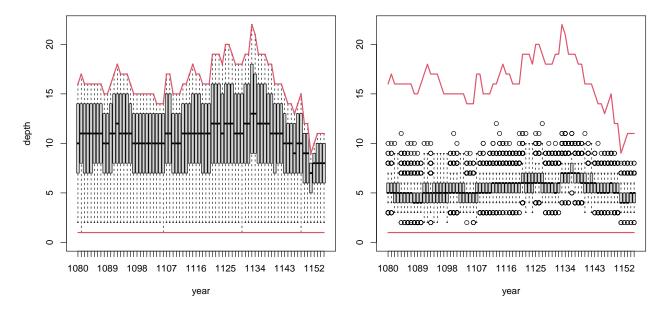
FIG 18. *(Left) Prior and (Right) posterior depth distributions from the analysis in Sec. 6.3 with $K = 11$. The x-axis shows years. Each box represents the posterior distribution for its year. It is computed using prior samples (at Left) and MCMC posterior sample depths (at Right). The upper and lower bounding lines indicate the least (equal one) and greatest (equal m_t in year t) depth possible in a given year.*
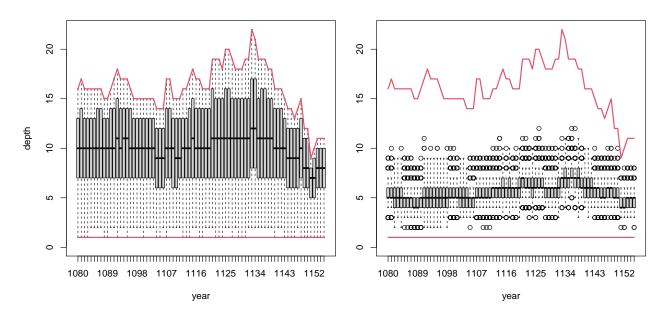

FIG 19. *(Left) Prior and (Right) posterior depth distributions from the analysis in Supp G, as Fig 18, here $K = 22$.*
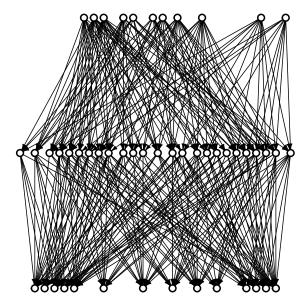
FIG 20. *A poset on 50 nodes drawn approximately uniformly at random from $\mathcal{H}_{[50]}$ using the MCMC algorithm given in* Muir Watt (2015)*, illustrating the concentration on orders of depth three.*

APPENDIX E: MCMC

Proposals for ρ and θ (which are in $[0,1]$) are prior draws with a Hastings ratio which does not involve the final likelihood factor in Eqn 17. Candidate states $\beta_r \to \beta'_r$, $r = 1, ..., S$ for the components of $\beta \in \mathcal{B}_0$ or $\mathcal{B}_S$ are simple random walk proposals with a fixed bandwidth (rejecting if the order condition in $\mathcal{B}_S$ is violated).

Proposals for the $\text{VAR}_{K,\rho,\theta}^{(b_j,e_j)}(1)$ time series U for each component $j \in \mathcal{M}_t$ target each time $t \in [b_j, e_j]$ in turn. The candidate state $\tilde{U}_j^{(t)} \in \mathbb{R}^K$ is a draw from the prior conditioned on ρ and θ and the values U_j^{t-1}, U_j^{t+1} of the process before and after the target vector, so that

$$\tilde{U}_j^{(t)} \sim N \left(\frac{\theta}{1+\theta^2}(U_j^{(t-1)} + U_j^{(t+1)}), \frac{1-\theta^2}{1+\theta^2}\Sigma^{(\rho)} \right), \qquad \text{if } b_j < t < e_j,$$

$$\tilde{U}_j^{(b_j)} \sim N \left(\theta U_j^{(b_j+1)}, (1-\theta^2)\Sigma^{(\rho)} \right), \qquad \text{if } t = b_j < e_j,$$

$$\tilde{U}_j^{(e_j)} \sim N \left(\theta U_j^{(e_j-1)}, (1-\theta^2)\Sigma^{(\rho)} \right) \qquad \text{if } b_j < t = e_j,$$

$$\tilde{U}_j^{(t)} \sim N \left(0_K, \Sigma^{(\rho)} \right) \qquad \text{if } t = b_j = e_j$$

with special cases at the start and end of the active period for bishop j. This determines the candidate state $\tilde{U}$ (which differs from U at just a single vector $U_j^{(t)} \to \tilde{U}_j^{(t)}$).

At U and β updates we update $\tilde{Z} = Z(\tilde{U}, \beta; s)$ and $\tilde{Z} = Z(U, \tilde{\beta}; s)$ respectively to determine the candidate poset $\tilde{h} = h(\tilde{Z})$. In a U-update at year t, the poset series h and $\tilde{h}$ are equal except for the replacement $h^{(t)} \to \tilde{h}^{(t)}$ so cancellations leave the Hastings ratio depending on the likelihood for lists $\mathcal{I}_t(\tau) = \{i \in \mathcal{I} : \tau_i = t\}$ only. The acceptance probability is

$$\alpha(\tilde{U}^{(t)}|U^{(t)}) = \min \left\{ 1, \prod_{i \in \mathcal{I}_t(\tau)} \frac{p(y_i|\tilde{h}^{(t)}[o_i], p)}{p(y_i|h^{(t)}[o_i], p)} \right\}.$$

Updates to components of β affect Z over all times so the full likelihood ratio is needed.

Proposals for τ are prior draws. The Hastings ratio depends only on likelihood factors for the list $i \in \mathcal{I}$ for which the date τ_i is being updated. If $\tilde{\tau}_i$ is the candidate value then the acceptance probability is

$$\alpha(\tilde{\tau}_i|\tau_i) = \min \left\{ 1, \frac{p(y_i|h^{(\tilde{\tau}_i)}[o_i], p)}{p(y_i|h^{(\tau_i)}[o_i], p)} \right\}$$

Proposals for p are also prior draws. This time the Hastings ratio is the full likelihood ratio, $p(y|h, \tau, \tilde{p})/p(y|h, \tau, p)$.

MCMC runs were initialised at either disordered or ordered initialisations. In the disordered initialisation all parameters were given starting values as far as possible from equilibrium values, the effects $\beta = 0_S$, the times τ are prior draws subject to $\tau \in [B, E]^N$ and the U-matrices specified so that the order was empty or near empty. In the ordered initialisation parameters start at values that might be typical for equilibrium, the effects β and the lists times τ as for disordered, and the U-matrices specified so that the initial posets $h^{(t,0)}$ are prior draws which dont conflict the data. This gives much deeper posets for the ordered initialisation. We checked that two runs started in ordered and disordered initialisations gave the same posterior densities for marginal parameter posteriors. We checked effective sample sizes and inspected MCMC traces.

APPENDIX F: POSTERIOR SUMMARY STATISTICS

F.1. Consensus Partial Order. We now define our principle posterior summaries. We seek a point estimate of the unknown true evolving status hierarchy of posets h. The consensus poset with threshold $\xi \in [0,1]$ displays all the relations supported at time t with estimated posterior probability

$$\hat{\xi}^{(t)}_{\langle j_1, j_2 \rangle} = L^{-1} \sum_{l=1}^{L} \mathbb{I}_{\langle i,j \rangle \in h^{(t,l)}} \tag{25}$$

greater than or equal ξ. It is a directed graph

$$\bar{h}^{(t)}(\xi) = (E_t(\xi), \mathcal{M}_t) \tag{26}$$

on the active bishops in year t with edges

$$E_t(\xi) = \left\{ \langle j_1, j_2 \rangle \in \mathcal{M}_t \times \mathcal{M}_t : \hat{\xi}^{(t)}_{\langle j_1, j_2 \rangle} \geq \xi \right\}.$$

The consensus poset need not be acyclic, even when $\xi \geq 0.5$, so $\bar{h}^{(t)}(\xi) \notin \mathcal{H}_t$ is possible. This seems to be possible but improbable: all the consensus partial orders in this paper are partial orders.

F.2. Testing for an effect due to seniority. In Sec. 6.2 we made an analysis of the full data with no constraint on the seniority effects, so $\beta \in \mathcal{B}_0$. Here we test for $\beta \in \mathcal{B}_8$.

We estimate a Bayes factor for the seniority effects β_r, $r = 1, ..., S$ to be decreasing with decreasing seniority (so increasing seniority-rank r). Let $\mathcal{B}_{S'} = \{\beta \in \mathcal{B}_0 : \beta_1 > \beta_2 > ... > \beta_{S'}\}$ be the event that the first $S' \leq S$ effects are ordered. It is feasible to compare models with $\beta \in \mathcal{B}_0$ against $\beta \in \mathcal{B}_{S'}$ for S' up to about $S' = 8$ using a Savage-Dickey estimator, as $\mathcal{B}_{S'} \subset \mathcal{B}_0$. We have

$$\begin{aligned} B_{S',0} &= \frac{p(y|\beta \in \mathcal{B}_{S'})}{p(y|\beta \in \mathcal{B}_0)} \\[6pt] &= \frac{\pi(\beta \in \mathcal{B}_{S'}|y)}{\pi(\beta \in \mathcal{B}_{S'})} \\[6pt] &= S'! \, E_{\rho, \theta, U, \beta, \tau, p|y}(\mathbb{I}_{\beta \in \mathcal{B}_{S'}}), \end{aligned} \tag{27}$$

as the marginals $\pi(\beta \in \mathcal{B}_0|y) = \pi(\beta \in \mathcal{B}_0) = 1$ when there is no constraint, and $\pi(\beta \in \mathcal{B}_{S'}) = 1/S'!$. The expectation can be estimated using samples from the unconstrained posterior $\pi(\rho, \theta, U, \beta, \tau, p|y)$. The point here is that the prior probability for $\beta \in \mathcal{B}_{S'}$ is very small for S' at all large, so if we see any posterior samples with $\beta \in \mathcal{B}_{S'}$ this shows the data is weighting strongly in favour of these configurations. However, even though we may have $\pi(\beta \in \mathcal{B}_{S'}|y) \gg \pi(\beta \in \mathcal{B}_{S'})$ the numerator in Eqn 27 is very small for large S' and hard to estimate accurately, so we restrict discussion to $S' \leq 8$.

We carried out this test using the setup in Sec. 6.2 (with $K = 11$ and likelihood $p_{(U)}$, so the model which gives the marginal β-distributions in the right plot in Fig 11). We find $\hat{B}_{6,0} = 4.0(1)$, $\hat{B}_{7,0} = 10(5)$ and $\hat{B}_{8,0} = 40(28)$ with standard errors from iid Bernoulli sampling in parenthesis (long ordered sequences are rare events in the MCMC, so effectively iid). The last of these is based on two MCMC samples with $\beta \in \mathcal{B}_8$. However, given the improbability of these states in the prior $(1/8!)$, seeing any samples at all with this pattern is evidence for decreasing seniority effects with decreasing seniority rank.
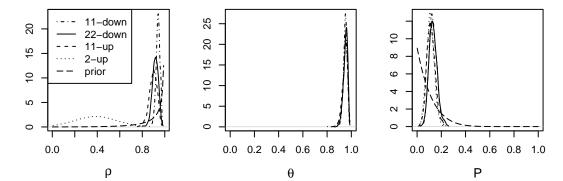
FIG 21. *Posterior parameter densities for ρ, θ and p from a selection of different models for the same data in the analysis of Sec. 6.2. The priors are given in Sec. 5.2 with $\beta \in \mathcal{B}_0$ and variations: (dot-dashed) $K = 11$, likelihood $p_{(D)}$ in Eqn 6; (dashed) $K = 11$, likelihood $p_{(U)}$ in Eqn 5; (solid) $K = 22$, likelihood $p_{(D)}$; (dotted) $K = 2$, likelihood $p_{(D)}$; (long-dashed, in ρ and p graphs) prior. The prior for θ is uniform.*

APPENDIX G: FURTHER RESULTS AND COMPARISON ANALYSES

G.1. Additional figures for the analysis in Sec. 6.2. Fig 22 shows MCMC traces for selected parameters and functions in the MCMC for the analysis in Sec. 6.2 using the MCMC algorithm of Supp E. This paper contains results from dozens of MCMC runs so we present this as specimen output. Another specimen set of MCMC traces are given in Fig 32 in Supp K.1 where we analyse synthetic data with the same model and structure as in this section (obtaining good recovery of truth). For the main analyses in Sec. 6.2 and 6.3 we ran two independent MCMC chains with ordered and unordered start states (see paragraph end of Supp E) and obtained excellent agreement.

Fig 21 shows comparisons of posterior distributions for the parameters ρ, θ and p for the unconstrained seniority-effects model $\beta \in \mathcal{B}_0$ (so, as Sec. 6.2) across runs with different model settings. The prior for each parameter is also plotted (except θ which is uniform). We vary the likelihood noise model (p_U and $p_{(D)}$) and the number of columns K in the status matrix $Z^{(t)}$. The serial-correlation parameter θ and the queue-jumping error probability p are almost unchanged when these other aspects of the model are varied.

The ρ posterior for $K = 2, 11$ and 22 (left panel of Fig 21) shift to larger values as K increases. Why does this happen? We can think of the posets h as fairly well informed by the data. In this setting the posets h restrict the possible U-matrices as the paths in U have to generate the posets in h. The U matrices function as "data" for the latent parameters ρ and θ. When we increase K the posets, informed by the data, are not changing (very much). The paths in U are longer at larger K but they must have the same patterns of intersection to give the same (or more or less the same) paths U and hence the same posets. When paths are longer the hazard for them crossing is greater. However, the fit can be maintained by making them flatter: each path must be more strongly correlated at large K than at small K. We conclude that there is some degree of non-identifiability between K and ρ.

G.2. Figures from the analysis in Sec. 6.3. In this section we show some further results from the analysis presented in Sec. 6.3, the analysis with constrained seniority effect parameters $\beta \in \mathcal{B}_S$ and $K = 11$ columns in the status matrix $Z^{(t)}$.

In Fig 23 we show transitive closures of the consensus poset $\bar{h}^{(t)}(\xi)$ defined in Eqn 26 for the years $t \in [1134, 1136]$. Edges with thresholds $\xi = 0.5$ and $\xi = 0.9$ are shown dashed and solid respectively. The reductions are shown in Fig 13. These transitively closed consensus
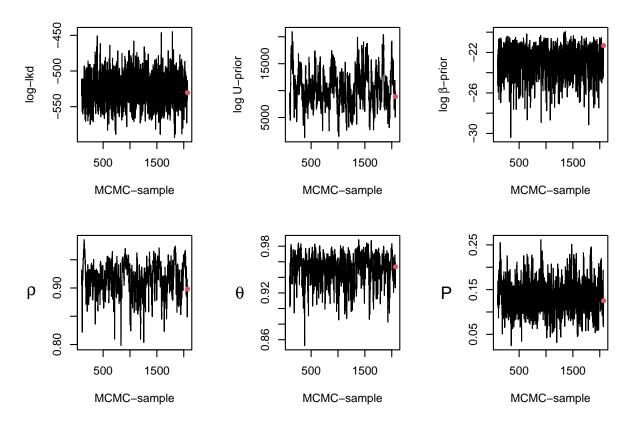
FIG 22. *Selected MCMC traces from the unconstrained seniority-effect analysis in Sec. 6.2.*

posets are hard to read so we prefer reductions. We show them to emphasise that many well-supported relations (solid edges) are found, as a sequence of weakly supported relations (dashed edges $\langle j_1, j_2 \rangle, \langle j_2, j_3 \rangle$) in the reduction tends to yield a well supported relation (a solid edge $\langle j_1, j_3 \rangle$) in the closure.

Fig 24 gives the mapping from names to numbers in all graphs in this paper.

Fig 25 shows the posterior mean status curves $\bar{Z}_j^{(t)}$, $t \in [b_j, e_j]$ for each bishop $j \in \mathcal{M}$. This is defined as for $\bar{U}_j^{(t)}$, $t \in [b_j, e_j]$ in Eqn 20. This figure should be compared with Fig 14. For further discussion of this pair of figures see Sec. 6.3.

Fig 26 shows 90% HPD sets for the posterior distribution of the list times τ_i, $i \in \mathcal{I}$. The x-axis in this graph is the list index (sorted, so that prior mean list times increase). Recall the data gives us a time range $\tau_i \in [\tau_i^-, \tau_i^+]$ for each list. The red bars show this time interval constraint, and also represent the prior, which is uniform. The dates of some lists are known without uncertainty (so $\tau_i^- = \tau_i^+$). These lists have been omitted from Fig 26. It can be seen that the data do inform the list times a little (the grey set is restricted relative to the red bar) though the benefit is seen mainly in lists with substantial prior time ranges (the longer red bars). We learn little here as the prior constraints $\tau_i \in [\tau_i^-, \tau_i^+]$ are already quite tight. The principle purpose of including the uncertain list dates in the analysis is to feed through the uncertainty in list dates into the estimates of average authority $\bar{U}_j^{(t)}$ and consensus partial orders $\bar{h}^{(t)}(\xi)$ rather than to do list dating.
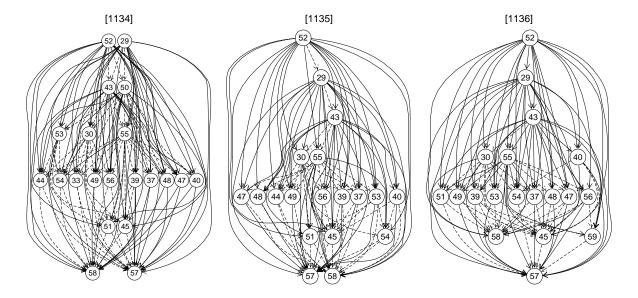
FIG 23. *Consensus partial orders for the years 1134-1136 (a selection from the years 1080-1155) under the constrained-effects model of Sec. 6.3. These are the transitive closures corresponding to the reductions in Fig 13 in Sec. 6.3. Dashed edges have posterior support at least $\xi = 0.5$. Black edges have support above 0.9. Vertex numbering as Fig 12.*

1) Wulfstan, bishop of Worcester

2) Walkelin, bishop of Winchester, 1070–1198

3) Robert, bishop of Seez

4) Osmund, bishop of Salisbury

5) Gundulf, bishop of Rochester

6) Gilbert, bishop of Lisieux

7) Remigius, bishop of Lincoln

8) Robert, Losinga, bishop of Hereford

9) Gilbert, bishop of Evreux

10) Geoffrey, bishop of Coutances

11) Stigand, bishop of Chichester

12) Peter, bishop of Chester

13) Odo, bishop of Bayeux

14) Michael, bishop of Avranches

15) Arfastus, bishop of Thetford

16) Osbern, bishop of Exeter

17) William, de Saint–Calais, bishop of Durham

18) Maurice, bishop of London

19) Robert, de Limesey, bishop of Chester

20) John, bishop of Bath

21) Ralph, bishop of Chichester

22) Hervey, bishop of Bangor

23) Robert, Bloet, bishop of Lincoln

24) Turgis, bishop of Avranches

25) Samson, bishop of Worcester

26) Turold, de Envermeu, bishop of Bayeux

27) Ranulf, Flambard, bishop of Durham

28) William, Giffard, bishop of Winchester, 1100–1129

29) Roger, Bishop of Salisbury

30) John, Bishop of Lisieux

31) Reinhelm, bishop of Hereford

32) Richard, Bishop of Bayeux

33) Urban, bishop of Llandaff

34) Richard, de Belmeis I, bishop of London, 1108–1127

35) Roger, bishop of Coutances

36) Theulf, Bishop of Worcester

37) Ouen, Bishop of Evreux

38) Arnulf, bishop of Rochester

39) Bernard, Bishop of St David's

40) Everard, bishop of Norwich

41) Richard, de Capella, bishop of Hereford

42) Robert, Peche, bishop of Chester, 1121–6

43) Alexander, Bishop of Lincoln

44) Godfrey, bishop of Bath

45) John, Bishop of Sees

46) Richard, bishop of Coutances

47) Simon, Bishop of Worcester

48) John, Bishop of Rochester

49) Seffrid, Bishop of Chichester

50) Gilbert, the Universal, bishop of London

51) Roger, de Clinton, Bishop of Chester

52) Henry, de Blois, Bishop of Winchester, 1129–1171

53) Robert, de Bethune, Bishop of Hereford

54) Algar, Bishop of Coutances

55) Nigel, Bishop of Ely

56) Geoffrey, Rufus, Bishop of Durham

57) Adelulf, Bishop of Carlisle

58) Richard, de Beaufeu, Bishop of Avranches

59) Robert, of Lewes, Bishop of Bath

60) Robert, de Sigillo, Bishop of London

61) Jocelin, de Bohun, bishop of Salisbury

62) William, Turbe, Bishop of Norwich

63) Hilary, Bishop of Chichester 1146–1169

64) Walter, Bishop of Rochester

65) Robert, de Chesney, Bishop of Lincoln

66) Richard, de Belmeis II, bishop of London 1152–1162
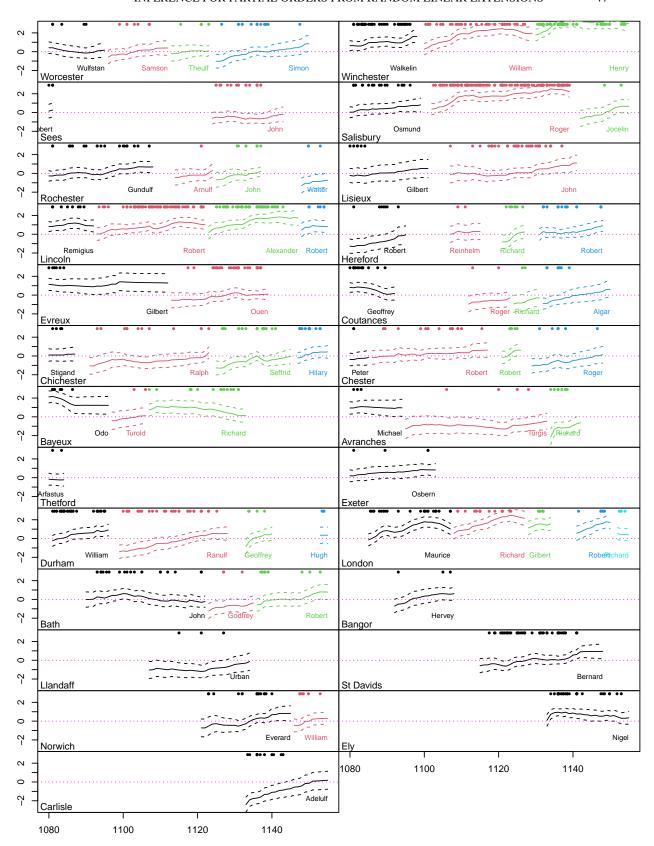
67) Hugh, du Puiset, Bishop of Durham

FIG 24. *Name index for vertices in the order graphs in Figs 1, 12, 13, 23 and 27.*

FIG 25. *Bishop-status curves* $\hat{Z}_j^{(t)}$ *(y-axis values, solid curves) plotted for each bishop* $j \in \mathcal{M}$ *as a function of time (x-axis, the year) from* b_j *to* e_j *with standard errors at one-sigma. The dots along the top of each graph show the times of the lists in which the color-matched bishop below appeared. This is output from the constrained seniority effects analysis in Sec. 6.3.*
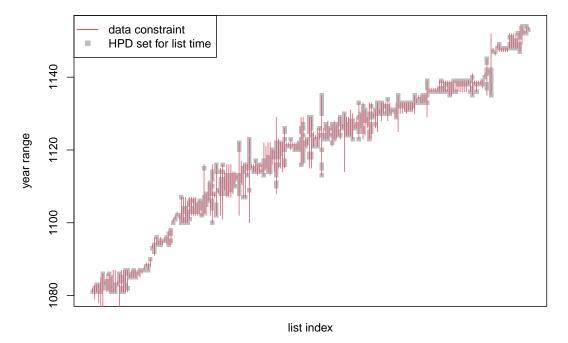
FIG 26. *Posterior distributions of the unknown list times* τ_i, $i \in \mathcal{I}$. *Some lists have known times, shown are the posteriors for uncertain list times. The x-axis is the list index (sorted by prior mean time) and the y-axis is in years CE. Each thin vertical line shows the known constrained interval* $[\tau_i^-, \tau_i^+]$ *for each list. The grey region over each vertical line is a 90% HPD set for the estimated age of the associated list. This is output from the constrained seniority effects analysis in Sec. 6.3.*

Finally we plot the full set of consensus partial orders for the analysis in this section. These are most easily viewed in the online version of this paper using a zoom function. The graphs for the years 1134-1136 were selected and plotted in Fig 13 for ease of viewing.
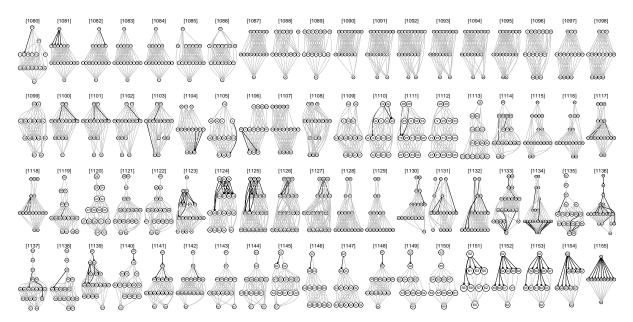
FIG 27. *Consensus partial orders for all years 1080-1155 under the constrained-effects model of Sec. 6.3. These are transitive reductions. Grey edges have posterior support at least $\xi = 0.5$. Black/solid edges have support above 0.9. Vertex numbering as Fig 12.*

G.3. Results from analyses like Sec. 6.3 but with $K = 2$ and $K = 22$. Here we gather together results from analyses identical to that in Sec. 6.3 (constrained seniority effects, $\beta \in \mathcal{B}_S$) but with $K = 11$ there replaced by $K = 2$ and then $K = 22$. We looked at $K = 2$ and $K = 22$ for the unconstrained analyses in Supp G.1 and considered how our estimates for ρ, θ and p were affected. When we make a constrained analysis we are aiming to produce final results to pass onto historians and in particular, relations need to be well-estimated.

Changing the number of features/columns in the authority vector $U_j^{(t)}$ changes the prior distribution over posets. The depth distributions are similar, but there is higher weight on lower depth posets at larger K. This can be seen by comparing the two graphs in Fig 17. The effect is visible in Fig 18 where the "whiskers" in the boxplot for $K = 22$ often reach from the maximum depth to the minimum depth, which is rare in the $K = 11$ case.

It may be desirable to use the $K = 22$ prior as the $K = 11$ prior puts very little probability on posets of depth one (not zero, but small). Depth one corresponds to the empty order. While putting low prior probability on the empty order reflects prior knowledge (it is very unrepresentative of what we know about the actors in our data), it is a little stronger than is really desirable, so some sensitivity analysis is in order.

In contrast the $K = 2$ prior weights quite strongly in favor of deeper partial orders. This might be improved by some compensating adjustment of the ρ−prior. However, for the purpose of this comparison we simply changed K and left all other aspects of the model unchanged. If we could "get away" with using $K = 2$ we certainly would, as the dimension of the authority parameters U is $\dim(U) = K \times \sum_t m_t$. In Sec. 6.3, where $K = 11$ we fit with $\dim(U) = 13,453$. This drops to $\dim(U) = 2446$ with $K = 2$ and makes for easier MCMC sampling. The class of two dimensional posets is still a very rich model space (it includes all VSPs). Also, at least for short time intervals, VSPs fit fairly well, with a Bayes Factor against the full poset model which is only weakly in favor of posets. VSPs have dimension at most

K	% relations	MAD
2	95%	0.02
11	100%	0
22	97%	0.01

TABLE 1

Results from comparison of posterior distributions at $K = 2$, 11 and 22 for the constrained analysis of Sec. 6.3 but varying the dimension K of the latent authority vectors $U_{j,:}^{(t)}$, $j \in \mathcal{M}$, $t \in [B, E]$. Comparisons are against $K = 11$ so it shares 100% of its relations with itself. The % relations column gives the percentage of agreeing estimated relations (including "no-relation") between the consensus posets for the two analyses in each comparison. The "MAD" column gives the Median Absolute Difference in the estimated probability for relations. Reported numbers are accurate to at least the reported significant digits.

two, so all VSPs can be represented by taking $K = 2$. This suggests that quite small values of K may sometimes be adequate.

Results are given in Table 1. We find conclusions are very similar for the three K-values. At the $\xi = 0.5$ threshold for consensus posets, 95% of estimated relations are the same between $K = 2$ and $K = 11$. However, the two analyses may support a relation at levels just above and below the consensus poset threshold, so this doesn't tell the whole story. Drilling down, if $i, j \in \mathcal{M}_t$ are two bishops active at time t and $\hat{\xi}_{\langle j_1, j_2 \rangle}^{(K,t)}$ (see Eqn 25 in Supp F.1) is the estimated posterior probability for the relation $j_1 \succ_{h^{(t)}} j_2$ in an analysis with $K \in \{2, 11, 22\}$ then the number of linearly independent $\hat{\xi}_{\langle j_1, j_2 \rangle}^{(K,t)}$-values we estimate in each analysis is 18912. The Median Absolute Difference (MAD) $|\hat{\xi}_{\langle j_1, j_2 \rangle}^{(11,t)} - \hat{\xi}_{\langle j_1, j_2 \rangle}^{(2,t)}|$ in these estimated probabilities is 0.02. Going from $K = 2$ to $K = 11$ isn't changing conclusions a great deal. However, if we do the same analysis comparing $K = 11$ and $K = 22$ we get slightly better agreement: 97% of relations agree and the MAD, $|\hat{\xi}_{\langle j_1, j_2 \rangle}^{(22,t)} - \hat{\xi}_{\langle j_1, j_2 \rangle}^{(11,t)}|$, is 0.01. We interpret these results to mean that once K is "large enough" we estimate the same relations.

APPENDIX H: PARAMETER ESTIMATES FROM MULTIPLE SHORT
TIME-WINDOWS

In the poset models of Secs. 6.2, 6.3 and Supp G.3 we assume that p, ρ, θ do not vary over time. One simple way to check this is to estimate their posterior distributions over short time windows. We use the same intervals (first column of Table 2) as Secs. 7.3 and J.2. Results are shown in Fig 28.
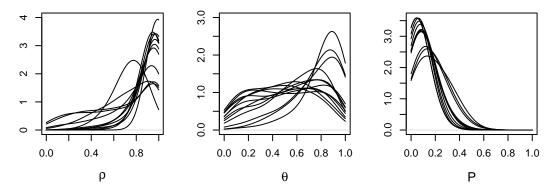

FIG 28. *Posterior distributions for ρ, θ and p estimated in each of the twelve time intervals listed in column 1 of Table 2 and discussed in Supp H*

The distributions for ρ and p show very little variation over time. The distributions for θ, the degree of correlation between the poset in one year and the next, do vary somewhat. This density is focused on large values for the first interval 1080-1084 and for intervals after about 1135. The other intervals are close to the (uniform) prior. However, this simply reflects the fact that we get little information about a correlation from a short time interval (5 years). We see no evidence here against our assumption that ρ, θ and p are constant in time.
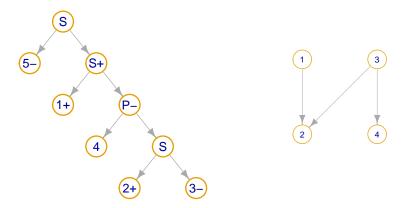
FIG 29. *(left) BDT representation of the poset in Fig 6. Working up from the bottom, 2 is above 3, 4 is unordered with respect to 2 and 3, 1 is above all three actors, and finally 5 is below all of 1-4. (right) The forbidden sub-graph F. A VSP order is any poset which does not contain a sub-graph isomorphic to F.*

APPENDIX I: VSP ORDER AND BUCKET ORDER ANALYSIS - FURTHER DETAILS

I.1. Further background on VSPs. Let $\mathcal{A}_1, \mathcal{A}_2$ be disjoint sets of actor labels and let $a_1 \in \mathcal{H}_{\mathcal{A}_1}$ and $a_2 \in \mathcal{H}_{\mathcal{A}_2}$ be posets. Let

$$a_1 \, \textcircled{S} \, a_2 = a_1 \cup a_2 \bigcup_{\substack{j_1 \in \mathcal{A}_1 \\ j_2 \in \mathcal{A}_2}} \{\langle j_1, j_2 \rangle\},$$

denote series combination setting all actors in a_1 "above" those in a_2 and let

$$a_1 \, \textcircled{P} \, a_2 = a_1 \cup a_2,$$

denote the parallel combination in which all actors in a_1 are unordered with respect to those in a_2. The class $\mathcal{V}_\mathcal{A}$ of VSP orders on $\mathcal{A}$ is defined recursively (Valdes et al., 1982): for $j \in \mathcal{A}$ let $\mathcal{V}_j = \{\emptyset\}$ to be the empty order on a single actor j; $\mathcal{V}_\mathcal{A}$ is the set of all posets H which can be decomposed as $H = a_1 \, \textcircled{S} \, a_2$ or $H = a_1 \, \textcircled{P} \, a_2$ for some partition $\mathcal{A}_1, \mathcal{A}_2$ of $\mathcal{A}$ and some $a_1 \in \mathcal{V}_{\mathcal{A}_1}$ and $a_2 \in \mathcal{V}_{\mathcal{A}_2}$.

A sequence of series or parallel combinations defining a given VSP is represented by a Binary Decomposition Tree (BDT, Valdes (1978)): leaves match elements of $\mathcal{A}$ and each internal vertex is an $\textcircled{S}$ or $\textcircled{P}$ operation on the VSPs rooted by its left and right child vertices. The BDT displayed in Fig 29 represents the poset in Fig 6, which happens to be a VSP. The "plus" and "minus" symbols on the vertices below an S-vertex indicate the upper and lower VSP in a $\textcircled{S}$-operation. Any given VSP may have many BDT representations (Jiang et al. (2023) counts them). If $H \in \mathcal{V}_\mathcal{A}$ is a VSP then $|\mathcal{L}[H]|$ can be computed in a time linear in $|\mathcal{A}|$ (Wells, 1971).

I.2. Testing partial orders against restricted orders. In this section we consider data overlapping a time-window $[t_1, t_2]$ by at least a half. Let

$$y^{(t_1, t_2)} = \left\{ y_i : i \in \mathcal{I}, \ \frac{1 + \min(t_2, \tau_i^+) - \max(t_1, \tau_i^-)}{1 + \tau_i^+ - \tau_i^-} \geq 0.5 \right\}.$$

denote this windowed data. Consider the Bayes factor

$$B_{\mathcal{V}, \mathcal{H}} = \frac{p(y^{(t_1, t_2)} | h \in \mathcal{V}^{(t_1, t_2)})}{p(y^{(t_1, t_2)} | h \in \mathcal{H}^{(t_1, t_2)})},$$

TABLE 2
VSP and Bucket Order Bayes Factors

Period	Bucket Orders					VSP orders				
	Prior	Post.	BF	ESS	BF std.err.	Prior	Post.	BF	ESS	BF std.err.
1080-1084	0.085	0	0	-	-	0.22	0	0	-	-
1086-1090	0.15	0.17	1.2	45	0.38	0.31	0.34	1.1	33	0.27
1092-1096	0.27	0.14	0.53	29	0.24	0.48	0.31	0.65	53	0.13
1104-1108	0.18	0.12	0.66	31	0.32	0.36	0.21	0.6	29	0.21
1110-1114	0.42	0.25	0.6	81	0.12	0.72	0.64	0.89	230	0.044
1118-1122	0.1	0.22	2.1	38	0.65	0.24	0.53	2.2	69	0.25
1126-1130	0.16	0.13	0.86	75	0.25	0.32	0.35	1.1	60	0.19
1128-1132	0.084	0.016	0.19	96	0.15	0.22	0.067	0.31	15	0.3
1132-1134	0.088	0.0053	0.06	950	0.027	0.22	0.045	0.21	230	0.063
1138-1142	0.13	0.024	0.18	80	0.13	0.28	0.11	0.37	69	0.13
1144-1148	0.4	0.59	1.5	99	0.12	0.68	0.87	1.3	330	0.027
1150-1154	0.2	0.17	0.83	32	0.32	0.39	0.36	0.93	45	0.19

with marginal likelihoods $p(y^{(t_1,t_2)}|h \in \mathcal{V}^{(t_1,t_2)})$ and $p(y^{(t_1,t_2)}|h \in \mathcal{H}^{(t_1,t_2)})$. Replace $\mathcal{V}$ with $\mathcal{K}$ to give $B_{\mathcal{K},\mathcal{H}}$ for Bucket Orders. Because the sets $\mathcal{K}^{(t_1,t_2)} \subset \mathcal{V}^{(t_1,t_2)} \subset \mathcal{H}^{(t_1,t_2)}$ are nested, these Bayes factors are easily estimated using Savage-Dickey estimators and MCMC samples from the full posterior $\pi(\rho,\theta,U,\beta,\tau,p|y^{(t_1,t_2)})$ in Eqn 17. We have

$$(28) \qquad B_{\mathcal{V},\mathcal{H}} = \frac{\pi(\mathcal{V}^{(t_1,t_2)}|y^{(t_1,t_2)})}{\pi(\mathcal{V}^{(t_1,t_2)})},$$

where

$$\pi(\mathcal{V}^{(t_1,t_2)}|y^{(t_1,t_2)}) = E_{\rho,\theta,U,\beta,\tau,p|y^{(t_1,t_2)}} \left(\mathbb{I}_{h(Z(U,\beta;s)) \in \mathcal{V}^{(t_1,t_2)}} \right)$$

is the posterior expectation and $\pi(h \in \mathcal{V}^{(t_1,t_2)})$ is the corresponding prior expectation.

I.3. VSP and Bucket Order Bayes Factor Estimates. The estimates used to make the plot in Fig 15 are given here. These numbers are discussed in Sec. 7.3. The effective sample sizes are for the MCMC output $\mathbb{I}_{h^{(t)} \in \mathcal{X}^{(t_1,t_2)}}$ where $\mathcal{X}$ is $\mathcal{B}$ or $\mathcal{V}$ and $h^{(t)} = h(Z(U^{(t)},\beta;s))$, $t \in [B,E]$ is taken from a chain targeting $\pi(\rho,\theta,U,\beta,\tau,p|y^{(t_1,t_2)})$ using the lists in the time interval $[t_1,t_2]$. These are the samples used to estimate the numerator in Eqn 28. The large ESS values in this table are associated with low posterior probabilities for the events $h \in \mathcal{B}^{(t_1,t_2)}$ and $h \in \mathcal{V}^{(t_1,t_2)}$. The resulting rare event process is approximately iid. It is unclear to us how to measure uncertainty reliably for these cases, but these Bayes factors must be small, as the prior probabilities for the same events are relatively large, and the posterior probabilities are at least much smaller.

APPENDIX J: PLACKETT-LUCE MODELS

In Supp J.1 and J.2 we define and fit two different types of Plackett-Luce model (Luce, 1959; Plackett, 1975). This is done for comparison purposes.

J.1. Comparison with results from a Plackett-Luce time-series model. Here we specify and fit a time-series Plackett-Luce model like that given in Glickman and Hennessy (2015). Our purpose is to test the conclusions we reached using the poset model and show that we get similar results when we use a qualitatively different model for the same data. We fit the Plackett-Luce time-series model to the full data from 1080-1155 and compare with results in Secs. 6.2 and 6.3. The model is simpler, there is no partial order, but it does share some parameters (β) and some parameters play very similar roles (λ and U). Results are remarkably similar. In the next section we will consider formal model comparisons.

J.1.1. *Parameters and Likelihood.* For $t \in [B, E]$ denote by $\lambda_j^{(t)}$ the authority of bishop $j \in \mathcal{M}$ in year t. Notation for seniority s and seniority effects β are unchanged from Sec. 3.2.2. The scalar time series $\lambda_j = (\lambda_j^{(t)})_{t=B}^E$ plays a similar role to the multidimensional time series of authority variables U_j defined in Sec. 3.2.3. However, for ease of coding we defined $\lambda_j^{(t)}$ for all $t \in [E, B]$ not just $t \in [b_j, e_j]$ as we did for U_j. This means $\lambda = (\lambda_j)_{j \in \mathcal{M}}$ is a "full" $M \times T$ matrix. Denote by $\lambda^{(t)} = (\lambda_j^{(t)})_{j \in \mathcal{M}}$ the column vector of λ-values in each time $t \in [B, E]$.

The times of some of the lists are uncertain so we fixed these to be the centres of their intervals, $\hat{\tau}_i = (\tau_i^- + \tau_i^+)/2$. The Plackett-Luce likelihood is then

$$(29) \qquad p_{PL}(y|\lambda, \beta) = \prod_{i=1}^N \prod_{j=1}^{n_i} \frac{\exp(\lambda_{y_{ij}}^{(\hat{\tau}_i)} + \beta_{s_{\hat{\tau}_i, y_{ij}}})}{\sum_{k=j}^{n_i} \exp(\lambda_{y_{ik}}^{(\hat{\tau}_i)} + \beta_{s_{\hat{\tau}_i, y_{ik}}})}$$

The "linear predictor" $\lambda_j^{(t)} + \beta_{s_{t,j}}$ captures the "status" of bishop j in year t just as the vector $Z_j^{(t)}$ did in Sec. 3.2.1. The covariates enter in the same way as before (Eqn 11) and so $\lambda_j^{(t)}$ is the part of status not determined by seniority.

J.1.2. *Plackett-Luce Priors.* The likelihood in Eqn 29 is invariant under $\lambda \to \lambda + \delta$ and $\beta \to \beta + \delta'$ for any $\delta, \delta' \in \mathbb{R}$. As noted above in Sec. 5.2.1, the U, β-latent variable model for posets has the same feature. We now remove this non-identifiability in our parameterisation.

Consider a projection matrix

$$Q_M = I_M - \frac{1}{M} \mathbf{1}_M \mathbf{1}_M^T,$$

with I_M the $M \times M$ identity matrix and $\mathbf{1}_M$ a column of M ones. Let Q_S be defined similarly. The matrix Q_M projects vectors in $\mathbb{R}^M$ to centred vectors $\{v \in \mathbb{R}^M : \sum_{j \in \mathcal{M}} v_j = 0_M\}$. Under a Plackett-Luce model with covariates for seniority, the generative model for authority is a centred $AR(1)$ process,

$$\lambda^{(t)} = \theta \lambda^{(t-1)} + Q_M \epsilon^{(t)},$$

where $\epsilon^{(t)} \sim N(\mathbf{0}_M, \sigma^2 I_M)$ iid for $t = B + 1, ..., E$. If $\lambda^{(t-1)}$ is centred, then $\lambda^{(t)}$ is also centred. The process is initialised in its prior equilibrium distribution,

$$\lambda^{(B)} = Q_M \epsilon^{(b)},$$

where

$$(30) \qquad \epsilon^{(b)} \sim N\left(0_M, \frac{\sigma^2}{1-\theta^2} I_M\right).$$

Since

$$(31) \qquad \lambda_1^{(t)} = -\sum_{j=2}^{M} \lambda_j^{(t)}, \quad \beta_1 = -\sum_{r=2}^{S} \beta_r,$$

we work with a non-singular prior density with parameters $\lambda_{2:M}^{(t)}$, $t \in [B, E]$. Denote by $Q_{2:M}$ the $M-1 \times M$ matrix obtained by dropping the first row from Q_M (with $Q_{2:S}$ defined similarly). The densities determined by these generative models are

$$\pi(\lambda_{2:M}^{(t)}|\lambda_{2:M}^{(t-1)}, \theta, \sigma) = N\left(\lambda_{2:M}^{(t)}; \theta\lambda_{2:M}^{(t-1)}, \sigma^2 Q_{2:M} Q_{2:M}^T\right),$$

$$\pi(\lambda_{2:M}^{(B)}|\theta, \sigma) = N\left(\lambda_{2:M}^{(B)}; \mathbf{0}_{2:M}, \frac{\sigma^2}{1-\theta^2} Q_{2:M} Q_{2:M}^T\right),$$

and

$$\pi(\lambda_{2:M}|\theta, \sigma) = \pi(\lambda_{2:M}^{(B)}|\theta, \sigma) \prod_{t=B+1}^{E} \pi(\lambda_{2:M}^{(t)}|\lambda_{2:M}^{(t-1)}, \theta, \sigma).$$

The prior on β is similarly

$$(32) \qquad \pi(\beta) = N(\beta_{2:S}; \mathbf{0}_{S-1}, Q_{2:S} Q_{2:S}^T).$$

There is nothing "special" about λ_1 or β_1 in these priors. They are determined by Eqns 31, but they are still exchangeable with the other parameters in the prior. We take a Unif$(0,1)$ prior for θ and a Gamma$(2,2)$ for σ as capturing reasonable variation.

J.1.3. *Posterior distribution.*   The Plackett-Luce posterior is then

$$(33) \qquad \pi_{PL}(\lambda_{2:M}, \beta_{2:S}, \sigma, \theta|y) \propto p_{PL}(y|\lambda, \beta)\pi(\lambda_{2:M}|\sigma, \theta)\pi(\beta_{2:S}, \sigma, \theta),$$

with λ_1 and β_1 in λ and β appearing in $p_{PL}(y|\lambda, \beta)$ given by Eqns 31.

We extended the λ_j-process for each bishop $j \in \mathcal{M}$ outside the range $[b_j, e_j]$. This made coding easier, but this means a large set of extra parameters must be brought into equilibrium in the target. However, by the stationarity of an $AR(1)$ process, the posterior marginal (integrating over $\lambda_j^{(t)}$ for $t \notin [b_j, e_j]$) of the fitted model is identical to the model which "starts" the λ_j-process at $t = b_j$ with the $AR(1)$ equilibrium in Eqn 30 and terminates it at $t = e_j$.

J.1.4. *Results for Plackett-Luce Time-series model.*   We use a simple random-walk MCMC algorithm to target $\pi_{PL}(\lambda_{2:M}, \beta_{2:S}, \sigma, \theta|y)$, on the full data of Secs. 6.2 and 6.3. We took unconstrained $\beta \in \mathcal{B}_0$. Denote by $\lambda_j^{(t,l)}$, $t \in [B, E]$, $j \in \mathcal{M}_t, \theta^{(l)}, \sigma^{(l)}$, $l = 1, ..., L$ the realised MCMC samples after thinning and burn-in. The Effective Sample Sizes for the parameters were $\theta(41)$, $\sigma(23)$, $\beta(21 \text{ parameters}, 538 - 2016)$ $\lambda(5092 \text{ parameters}, 55 - 6779)$. Mixing for the key parameter vectors β and λ was fair, and the agreement between results from the two analyses in this section and Secs. 6.2 and 6.3 supports our conclusion that the samples are representative.

Results for quantities of historical interest are very similar to those obtained using the partial-order posterior in Eqn 17. The posterior means for θ and σ were $0.93(0.02)$ and
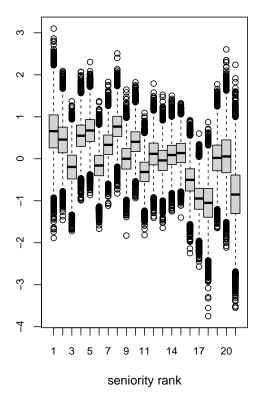
FIG 30. *(Left) Marginal posterior distributions of seniority effect parameters from the unconstrained seniority effects analysis in Sec. 6.2 with likelihood $p_{(D)}$ in Eqn 5. (Centre) As left with likelihood $p_{(U)}$ in Eqn 6. (Right) Corresponding distributions estimated using the Plackett-Luce time-series model in Supp J.1.*

$0.69(0.085)$ (posterior standard deviation in parenthesis). The posterior for θ is similar to its poset counterpart in Fig 10. Posterior effect β distributions are plotted in Fig 30. They are qualitatively similar to those for the poset-analyses in Fig 11. However, seniority effects in Fig 30 do not decline as smoothly as we know they should, and clearly not as well in line with historical expectations as the corresponding poset estimates in Fig 11. This suggests PL is using the degrees of freedom in the seniority effects to accommodate misspecification in the observation model. In Fig 31 we plot the curves

$$\hat{\lambda}_j^{(t)} = \frac{1}{L} \sum_{l=1}^{L} \lambda_j^{(t,l)}$$

over $t \in [b_j, e_j]$ for each j arranged by diocese as we did for $\bar{U}_j^{(t)}$ defined in Sec. 5.2.1. The resulting authority-measures show similar dependence on time. This shows that our results for these features are robust.

J.2. Comparison with a Plackett-Luce mixture model.

J.2.1. *Parameters and Model.* In this section we define a Plackett-Luce mixture model for list data taken from an interval of time $[t_1, t_2]$ which is short enough to allow us to drop time dependence and covariates. The Leave-One-Out cross validation we use below is time consuming and we do this for ease of computation. We fit this model and a poset model without covariates and carry out model comparison between the two models. Dropping time-varying covariates is natural when working in a short time window, as the effects due to bishop and seniority cannot be separated. Denote by $\mathcal{I}_{t_1,t_2}$ the indices of the
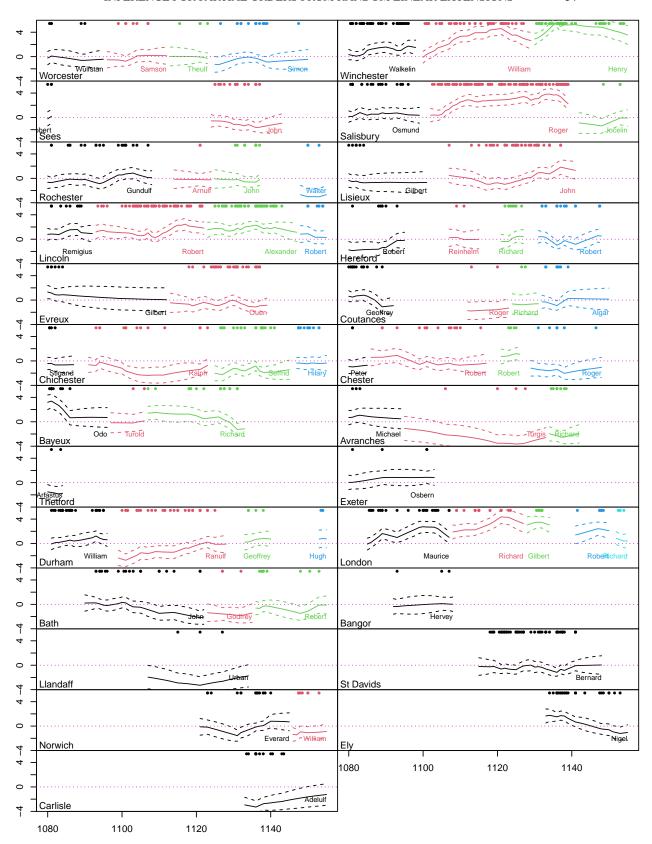
FIG 31. *Authority-curves* $\hat{\lambda}_j^{(t)}$ *(y-axis values) over* $t \in [b_j, e_j]$ *(x-axis, the year) for each bishop* $j \in \mathcal{M}$ *arranged by diocese, for the Plackett-Luce time-series model in Supp J.1. See Fig 14 for details and comparison.*

list data covered by the window (at least half of the years $[\tau_i^-, \tau_i^+]$ fall within $[t_1, t_2]$) and let $N_{t_1,t_2} = |\mathcal{I}_{t_1,t_2}|$ be the number of lists in this window. Let $y^{(t_1,t_2)} = (y_i)_{i \in \mathcal{I}_{t_1,t_2}}$ be the windowed data. A finite mixture of Plackett-Luce models was proposed as a robust model for ranked data with incomplete lists ("partially ranked data", Mollica and Tardella (2017, 2020). A mixture of D Plackett-Luce models for ranking a set $\mathcal{M}_{t_1,t_2} = \cup_{t \in [t_1, t_2]} \mathcal{M}_t$ containing $m_{t_1,t_2} = |\mathcal{M}_{t_1,t_2}|$ actors assumes that lists are sampled from a heterogeneous population composed of D sub-populations. Each sub-population is modeled by one of the mixture components with Plackett-Luce status vector $\lambda^{(d)} \in \mathbb{R}^{m_{t_1,t_2}}$, $d \in [D]$. Let $\lambda = (\lambda^{(d)})_{d \in [D]}$ be the $m_{t_1,t_2} \times D$ matrix of status parameters.

The likelihood for a mixture of D Plackett-Luce models is then

$$(34) \qquad p_{mix\,PL}(y^{(t_1,t_2)}|\lambda,\omega) = \prod_{i \in \mathcal{I}_{t_1,t_2}} \sum_{d=1}^{D} \omega_d \prod_{j=1}^{n_i} \frac{\exp(\lambda_{y_{i,j}}^{(d)})}{\sum_{k=j}^{n_i} \exp(\lambda_{y_{i,k}}^{(d)})},$$

where $\omega_1, ..., \omega_D$ are weights of mixture components and $\sum_{d=1}^{D} \omega_d = 1$. Noninformative priors suggested by Mollica and Tardella (2020) are assigned with $\exp(\lambda_j^{(d)}) \sim \text{Gamma}(1, 0, 001)$ for $j \in \mathcal{M}_{t_1,t_2}$ and $\omega_1, ..., \omega_D \sim Dir(1, ..., 1)$.

The posterior is

$$\pi_{mix\,PL}(\lambda, \omega) \propto \pi(\lambda, \omega) p_{mix\,PL}(y^{(t_1,t_2)}|\lambda, \omega).$$

We used the MCMC sampler available the R-package PLMIX provided by Mollica and Tardella (2020) to target $\pi_{mix\,PL}$. This sampler uses a data augmentation scheme due to Caron and Doucet (2012).

In order to fit the model we must select the number of mixture components D. This is related to Bayesian model comparison. The PLMIX offers several model selection criteria to choose the number of mixture components. We used the Deviation Information Criterion. The selected D-values are given in the third column of Table 3.

J.2.2. *Model comparison.* The expected log pointwise predictive density (ELPD, Vehtari et al. (2017)) measures the posterior predictive accuracy of a model. It is relatively straightforward to estimate and one natural measure of goodness of fit. Let $Y \in \mathcal{P}_O$ be a single generic list with membership $O \subseteq \mathcal{M}_{t_1,t_2}$. Suppose the true generative model for a list is $Y \sim f_O$. Denote by $p_{A,O}(Y|y_{y_1,y_2})$ the posterior predictive probability for Y in model A when the membership of list Y is O. The ELPD for model $A \in \{\text{Plackett-Luce, Partial order}\}$ conditioned on the observed data is

$$\text{elpd}(A|y^{(t_1,t_2)}) = \sum_{i=1}^{m_{t_1,t_2}} E_{Y \sim f_{o_i}}(\log p_{A,o_i}(Y|y^{(t_1,t_2)})),$$

$$= \sum_{i=1}^{m_{t_1,t_2}} \left[\sum_{Y \in \mathcal{P}_{o_i}} f_{o_i}(Y) \log p_{A,o_i}(Y|y^{(t_1,t_2)}) \right].$$

The expectation is (up to a constant) the negative of the KL divergence between the true generative model for a list with membership o_i and the model-A posterior predictive for that list. We favour models with larger values of the elpd.

We do not know f_O but we do know the data are sampled from this distribution, so we follow Vehtari et al. (2017) and use leave one out cross validation (LOO-CV) to estimate $\text{elpd}(A|y^{(t_1,t_2)})$. Our estimator is

$$(35) \qquad \widehat{\text{elpd}}(A|y^{(t_1,t_2)}) = \sum_{i=1}^{m_{t_1,t_2}} \log p(y_i|y_{(-i)}^{(t_1,t_2)})$$

where $y_{(-i)}^{(t_1,t_2)} = y^{(t_1,t_2)} \setminus \{y_i\}$ so the i-th observation is left out.

J.2.3. *Test results.*   In Table 3 the twelve short periods $[t_1, t_2]$ we chose are given in the first column. The same periods are used to for numerical comparison with VSP and Bucket Orders in Table 2. The number of lists m_{t_1,t_2} covered by each interval is in the second column. We estimate the posterior predictive probability for each omitted list in a period by averaging its likelihood over posterior samples, so the number of MCMC runs for each entry in columns four and five is equal to m_{t_1,t_2}. This impacted our choices for the locations of time intervals as we had to avoid periods with too few lists or too many long lists.

TABLE 3

ELPD estimates and standard errors for twelve periods.

Period	# lists	Plackett-Luce Mixture		Partial-order
$[t_1, t_2]$	$m_{[t_1,t_2]}$	components D	$\widehat{\text{elpd}}$	$\widehat{\text{elpd}}$
1080-1084	26	2	-162.52 (17.55)	-54.35 (11.80)
1086-1090	10	1	-61.43 (14.43)	-23.33 (6.26)
1092-1096	13	1	-49.46 (8.55)	-23.03 (7.98)
1104-1108	23	1	-95.56 (10.15)	-41.08 (9.69)
1110-1114	21	1	-59.93 (8.16)	-33.55 (9.62)
1118-1122	20	2	-89.01 (12.36)	-21.07 (6.99)
1126-1130	25	2	-113.03 (11.77)	-36.13 (9.21)
1128-1132	33	2	-190.34 (16.37)	-73.97 (13.30)
1132-1134	19	2	-110.49 (22.07)	-35.66 (9.91)
1138-1142	32	1	-186.09 (16.77)	-88.09 (25.58)
1144-1148	7	1	-42.15 (4.26)	-17.56 (8.33)
1150-1154	11	1	-60.05 (11.27)	-16.53 (4.20)

The elpd's in column five are uniformly larger than those in column four, so the poset model is clearly favoured over the Plackett-Luce mixture model for all these short periods. This may be because the poset is able to restrict the set of probable lists more tightly around the lists in the data, while the fitted Plackett-Luce mixture is relatively more dispersed over lists. The reported uncertainty estimate is just the sample variation of the mean of the m_{t_1,t_2}-samples used to form the estimate. It is likely to be an underestimate because the same lists appear in test and training positions in different terms in the sum, so its terms are correlated (Bengio and Grandvalet, 2004; Sivula et al., 2022).
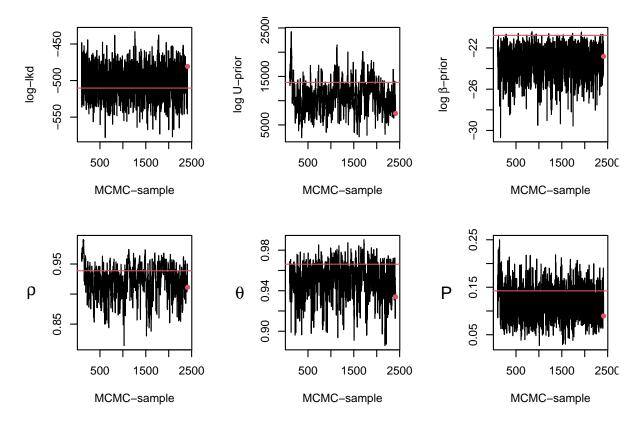
FIG 32. *Synthetic Data. Selected MCMC traces from the unconstrained seniority-effect analysis in Supp K.1. Horizontal line shows values at truth.*

APPENDIX K: CHECKS ON SYNTHETIC DATA

K.1. Checks on synthetic data: replicate main analysis. We simulated synthetic data with the same structure as the real data and checked we could recover true parameter values. We took true values of $\rho, \theta, \beta, U, \tau, p$ sampled from the unconstrained posterior with $K = 11$ discussed in Sec. 6.2 and simulated data $y'_i \sim p_{(U)}(\cdot | h^{(\tau_i)}[o_i], p), \ i \in \mathcal{I}$.

MCMC output targeting the posterior for the synthetic data y' is shown in Fig 32. The true ρ, θ and p values fall within the support of their respective posteriors in Fig 33. Credible intervals for $\bar{Z}$ in Fig 35 and β in Fig 34 cover the truth well.

In order to check reconstruction of relations in the fitted social hierarchy we report in Fig 36 the proportion of true relations which we would reject if we reject when the posterior odds for the relation are smaller than one third (the threshold for "positive" evidence in Kass and Raftery (1995)). This is equivalent to rejecting the relation $i \succ_{h^{(t)}} j$ for $i, j \in \mathcal{M}_t, \ t \in [B, E]$ if the estimated posterior probability $\hat{p}_{i,j}^{(t)}$ for the relation is less than one quarter. This runs at around 10% incorrect rejections at this level of evidence. We considered comparing the consensus partial order with the true partial order in each year and counting true positives and false negatives at a threshold $\xi = 0.5$. However, one has then to choose the threshold and, as in our discussion in Supp G.3, a relation with support just below a threshold of $\xi = 0.5$ is still the favored relation of the three possibilities (correct relation, relation but wrong order and no relation) so that would be misleading.
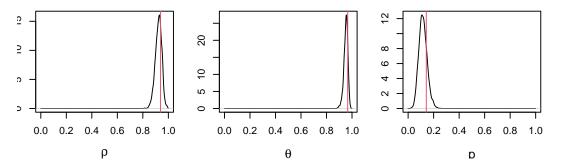
FIG 33. *Synthetic Data. Posterior parameter densities for ρ, θ and p from the unconstrained seniority effects analysis in Supp K.1. Vertical line shows true value.*
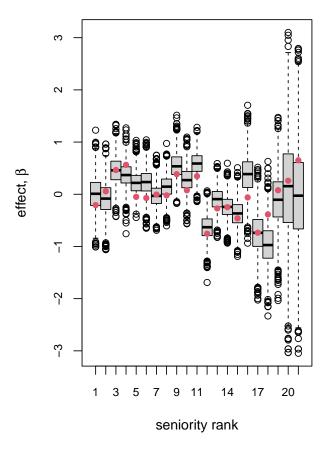

FIG 34. *Synthetic Data. Marginal posterior distributions of seniority effect parameters from the unconstrained seniority effects analysis in Supp K.1. Solid dots over boxes show true values.*

K.2. Checks on synthetic data: when the true orders are total orders. In some feedback from historians a concern was expressed that the relatively shallow posets we see reflect differences in how often bishops witness and that the "true" orders were deeper. We tested this by simulating synthetic data in which the true posets all had depth close to or equal the maximum m_t in each year $t \in [B, E]$, but using the same list memberships o_i, $i \in \mathcal{I}$ as the real data. For the purpose of this analysis we slightly reduced the data set, dropping some dioceses (Sherborne, Exeter, Bath, Thetford, Wells, Le Mans, Bangor, St Asaph, Tusculum,
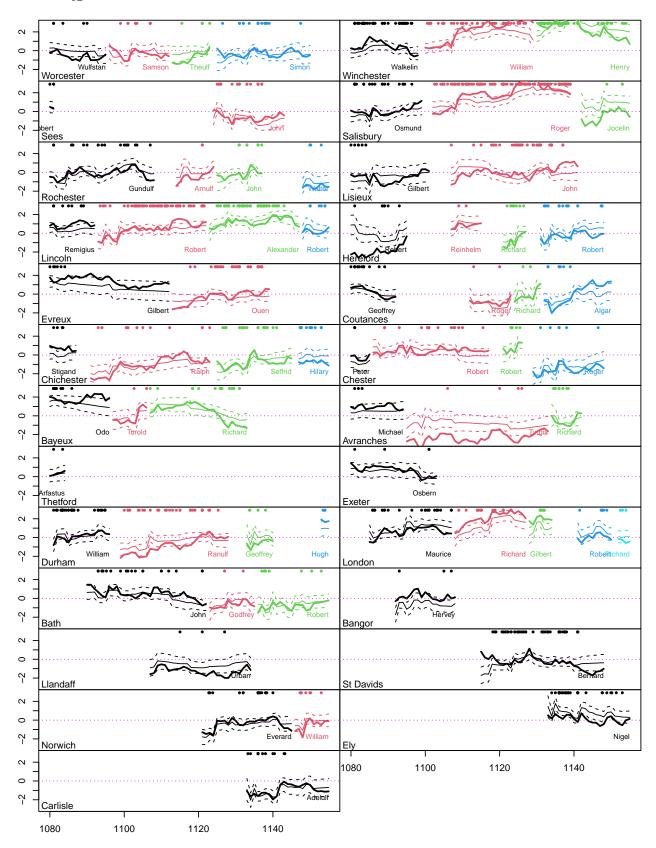
FIG 35. *Synthetic Data. Bishop-status curves $\hat{Z}_j^{(t)}$ (y-axis, thin solid curves) plotted for each bishop $j \in \mathcal{M}$ as a function of time (x-axis, the year) from b_j to e_j with standard errors at one-sigma. The dots along the top of each graph show the times of the lists in which the color-matched bishop below appeared. This is output from the constrained seniority effects analysis in Sec. K.1. The thick solid lines show the true values.*
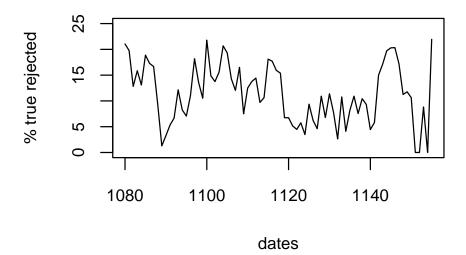
FIG 36. *Synthetic data analysis discussed in Supp K.1. The fraction of true relations rejected in the unconstrained seniority effects analysis of a synthetic data set simulated with $K = 22$ as a function of year.*

the Orkneys and Llandaff) that appear very infrequently in witness lists. This left $M = 59$ bishops and the same number $N = 371$ of lists.

We set the "true" noise at level $p^* = 0.1$ typical of that seen in our fit. We used the same time intervals for the lists as in the real data so the synthetic data has the same structure as the real data, but different list orders. We simulated a "true sequence" $h^* = h^{*(t)}$, $t \in [B, E]$ of high-depth orders (by setting $\rho^* = 0.99999$ close to one) and true observation times τ^* from the prior, then simulated synthetic-data lists according to the observation model $y_i' \sim p_{(U)}(\cdot | h^{*(\tau_i^*)}[o_i], p^*)$, $i \in \mathcal{I}$. We then simulated the posterior $\pi(\rho, \theta, U, \beta, \tau, p | y')$ and checked consensus partial orders and checked we recovered the true depth. The posterior depth distribution is shown in Fig 37 at right. The posterior shows good overlap with the truth despite it being on the boundary of the space.

In summary, we reconstructed the true total orders well with good depth. If the true orders were total orders we would see this in our analysis. We do not, so we conclude that the uneven list memberships, dates and list-lengths are not obscuring some hidden deeper order.

It is of interest to consider what happens when the true poset has depth 1, so the poset is the empty order and there are no constraints at all. The data y_i would then all be uniform random permutations of the list membership vectors o_i, $i \in \mathcal{I}$ and there is then a non-identifiability. The model can fit uniform random permutations by making $p \simeq 1$, so taking high noise, and any poset we like (as we get back uniform random permutations as $p \to 1$ irrespective of the underlying partial order) or by making the poset empty (so depth 1) and letting p be any value. In some experiments which we do not report on synthetic data with depth 1 we found that the model favoured the $p \to 1$ fit (as the prior weights against depth 1). Again, the fit for our data favours small noise, $p \simeq 0.1$ so by the same reasoning as for total orders, this non-identifiability is not an issue for the data we actually have. This has been called the "principle of sufficient reason": the prior is well behaved over the range of parameter values actually supported by the data.
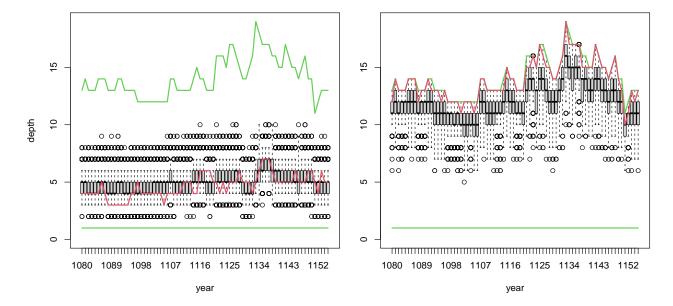
FIG 37. *Synthetic data taking the true order sequence h^* to be a sample from the Sec. 6.3-posterior (left, with the slightly reduced data set used in this appendix) or a sequence of deep orders sampled from the prior (right). Fitting with prior settings as Sec. 6.3 shows good overlap between truth and posterior. This suggests the prior is not biasing. The green lines above and below show the minimum and maximum possible poset depth in each year and the red line shows the depth of the true poset used to generate the synthetic data.*

APPENDIX L: FIXED-TIME MODEL

A simpler version of the model summarised in Sec. 3.3 outlined in Nicholls and Muir Watt (2011), without any time-series structure, may be of interest so we give the simplest model of this sort. Covariates depending on time (and their effects β) have been dropped. They might be replaced with covariates which vary across lists. However we do not pursue this.

In this fixed-time setting we have N lists with labels $i \in \mathcal{I}$, M bishops with labels $j \in \mathcal{M}$, Z is an $M \times K$ feature matrix (not a time series of matrices) with rows $Z_j \in \mathbb{R}^K$ giving the K status features for each bishop, and $H = h(Z)$ is a poset on $[M]$. All the lists have the same equal time, so τ is dropped. The generative model is

$$\rho \sim \text{Beta}(\gamma)$$
$$Z_j \sim N(0, \Sigma^{(\rho)}), \text{ iid for } j \in \mathcal{M}$$
$$p \sim \text{Beta}(1, \delta),$$

then $H = h(Z)$ using Eqn 10 and the data are realised

$$y_i \sim p(\cdot | H[o_i], p), \quad \text{independently for } i \in \mathcal{I}.$$

The joint posterior distribution is

$$(36) \qquad \pi(\rho, Z, p | y) \propto \pi(\rho, p) \prod_{j \in \mathcal{M}} N(Z_j; 0, \Sigma^{(\rho)}) \prod_{i=1}^{N} p(y_i | H[o_i], p).$$