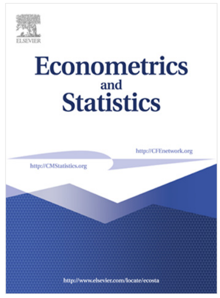
Journal Pre-proof



2nd Special Issue on Bayesian Methods in Statistics and Econometrics: Structured dependence and robust inference at scale

Roberto Casarin, Luciana DallaValle, Michele Guindani,
Daniel Kowal, Yasuhiro Omori

PII: S2452-3062(26)00009-2
DOI: https://doi.org/10.1016/j.ecosta.2026.03.001
Reference: ECOSTA 481

To appear in: *Econometrics and Statistics*

Please cite this article as: R. Casarin, L. DallaValle, M. Guindani et al., 2nd Special Issue on Bayesian Methods in Statistics and Econometrics: Structured dependence and robust inference at scale. *Econometrics and Statistics* (2026), doi: https://doi.org/10.1016/j.ecosta.2026.03.001.

2nd Special Issue on Bayesian Methods in Statistics and Econometrics: Structured dependence and robust inference at scale

This special issue of *Econometrics and Statistics* is devoted to Bayesian methods in statistics and econometrics. This collection can be seen as a timely reflection of the current state of Bayesian research at the interface of statistics and econometrics. Bayesian methodology has become central to the analysis of modern data not only because of its flexibility, but also because it can combine structural insight, computational advances, and careful uncertainty quantification within a single inferential framework. The papers assembled in this issue capture this evolution well, spanning problems driven by dependence, high dimensionality, and the growing complexity of contemporary applications.

This special issue originated from a call for papers inviting manuscripts that develop, illustrate, and showcase innovative Bayesian approaches across a broad range of domain areas (including environmental studies, biology and medicine, neurosciences, finance and economics, business and industry), with a dedicated space for Bayesian econometrics. The call emphasized a substantial *novel* Bayesian component and particularly encouraged contributions featuring developments in causal inference, (broadly defined) nonparametric methods, variable selection, and computationally efficient approaches to handling big data.

The contributions are organized around two closely connected developments that, in different ways, characterize much of contemporary Bayesian research. One is the search for structured and interpretable representations of dependence. The other is the construction of inferential procedures that remain reliable in the face of realistic data limitations and growing computational demands. The papers assembled here illustrate these directions across a broad range of models and applications: count time series motivated by large-scale digital traces, Bayesian nonparametric modeling for multivariate brain signals, causal inference under truncation because of death in longitudinal studies, flexible Bayesian testing for Granger causality, sparse factor modeling that confronts non-uniqueness and multi-modality, hierarchical Bayesian integration and clustering across multiple single-cell RNA-seq datasets, semi-supervised inference with explicit Bayesian debiasing, and high-dimensional, time-varying VAR forecasting via dynamic graphical modeling and fast variable selection.

Looking more closely at the contributions collected here, the first of these two strands highlights a particularly important role of Bayesian methodology: its ability to represent dependence in ways that are both flexible and interpretable. As an example, in count time-series analysis, apparent serial persistence and structural instability are often difficult to distinguish from artifacts induced by an overly rigid marginal model. Features such as excess zeros, asymmetry, heavy tails, or departures from equidispersion can easily be misread as evidence of stronger dependence or latent heterogeneity when the innovation distribution is too narrow to accommodate them directly.

Baltodano Lopez et al. [2025] expand the class of count time-series models by introducing a first-order integer-valued autoregressive process with Generalized Katz innovations, allowing a richer treatment of dispersion, asymmetry, and tail behavior in discrete dynamics. Their contribution not only enlarges the class of admissible count processes, but does so in a way that sharpens interpretation: by placing greater flexibility at the level of the innovations, their model helps disentangle true temporal dependence from distortions introduced by marginal mis-specification.

In a different but conceptually related direction, Granados-Garcia et al. [2024] propose a multivariate Bayesian nonparametric framework that represents high-dimensional multivariate time series as a low-dimensional mixture of latent oscillatory components. A distinctive design choice is the use of AR(2) spectral densities as unimodal kernels, enabling each latent component to be interpreted through a characteristic peak and bandwidth and facilitating connections to dependence summaries used in the signal-processing literature. The application to EEG data demonstrates the scientific value of this representation: it supports interpretable decompositions and association summaries while allowing the data to determine how many oscillatory components are needed and how they are shared across channels.

Gutiérrez et al. [2024] likewise focus on dependence in the form of directed temporal relationships, proposing a Bayesian nonparametric framework for Granger-causality testing in which the causal structure among time series is encoded through a latent binary matrix. This representation shifts attention away from separate coefficient-level tests and toward joint inference on the overall pattern of directed dependence. By combining a flexible mixture model with a prior over Granger-causal configurations, the paper shows how Bayesian methods can represent and learn Granger-causal structure in a way that is both flexible and inferentially coherent.

Kaufmann and Pape [2025] study dependence through sparse factor modeling, but with an important emphasis on non-uniqueness. Their central point is that a high-dimensional dataset may support more than one sparse factor loading matrix, with different sparse representations explaining roughly the same share of variation and therefore suggesting different interpretations of the underlying dependence structure. Their geometric analysis clarifies when such non-uniqueness can arise, particularly in the presence of pervasive factors together with weaker, more local ones, and their Bayesian procedures are explicitly designed to recover these alternative sparse structures. The contribution is therefore not only to sparse factor modeling itself, but to the interpretation of dependence in latent-variable models: it shows that the relevant inferential question may be not only which sparse structure is supported by the data, but whether that structure is unique in the first place.

Vahap [2025] represents dependence in high-dimensional vector autoregressions through a Bayesian dynamic graphical model in which both lagged and contemporaneous relationships are encoded by sparse directed graphs. The paper distinguishes between dynamic dependence, driven by autoregressive and cross-lagged effects, and contemporaneous dependence, driven by conditional relationships among variables within each time point, and uses pairwise conditional independence to translate both into locally structured regression problems. This decomposition makes the multivariate dependence structure not only computationally tractable, but also directly interpretable as an evolving graph of temporal and contemporaneous interactions.

The second strand running through this special issue emerges most clearly in settings where the inferential problem is made difficult not only by model complexity, but by the fact that the outcome of interest may itself become undefined for part of the population. Grossi et al. [2025] provide a particularly strong example of this challenge in their study of longitudinal causal inference under truncation by death. Their Bayesian principal stratification framework extends causal analysis to settings in which units may die at different times and outcomes are well defined only up to the corresponding moment of death. In such problems, the central difficulty is not simply missingness, but the need to define causal effects for latent subpopulations whose membership is only partially observed. By focusing on survivor-average causal effects for units that would survive up to a given time regardless of treatment assignment, the paper shows how Bayesian inference can be used to address problems in which careful treatment of latent structure, partial identification, and causal interpretability is indispensable.

If the previous paper highlights the need for robust Bayesian inference when outcomes may become undefined, Liu et al. [2024] show a complementary challenge arising from noisy, high-dimensional, and heterogeneous data observed across multiple datasets. In the single-cell RNA sequencing setting, they develop a hierarchical Bayesian model that jointly performs normalization, clustering, and dataset integration, thereby avoiding the fragility of multi-step workflows in which these tasks are separated. By using a hierarchical Dirichlet process, the model allows cell types to be shared across datasets while their proportions vary, and it clusters cells jointly through both mean expression and dispersion. The paper also directly addresses the computational burden of high-dimensional genomics by developing posterior computation strategies and consensus-style clustering summaries. The result is a framework that further exemplifies the second strand of the issue: it treats integration as an inferential problem that must preserve uncertainty and dataset-level heterogeneity, and it shows how hierarchical Bayes can provide robust borrowing of information.

Sert et al. [2025] bring the second strand into focus from a different angle, one in which the main challenge is not cross-dataset heterogeneity but the need to protect inference on the target parameter from the bias induced by complex nuisance estimation. In the semi-supervised setting they consider, unlabeled covariates are abundant but outcomes are observed only for a much smaller labeled sample, so that the potential gains from using the unlabeled data depend on learning a nuisance regression function that may be high-dimensional, slowly estimated, or even mis-specified. They develop a Bayesian debiased modeling and inference framework that targets the population mean by carefully chosing summary statistics rather than through a full model for the raw data, and that uses sample splitting and cross-fitting to separate nuisance learning from inference on the parameter of interest. Robust Bayesian inference here means not only handling noisy or heterogeneous data, but also ensuring that inference remains stable when the target depends on complex and potentially fragile nuisance estimation.

Several papers in the issue make the case (sometimes implicitly, sometimes explicitly) that computation is not an implementation detail but part of the inferential contribution. In the Bayesian dynamic graphical model of Vahap [2025], fast variable selection and recursion strategies are what allow time variation and volatility discounting to remain feasible in high-dimensional VARs. In the single-cell integration work, computational attention is essential for making hierarchical modeling practical at modern scRNA-seq scale [Liu et al., 2024]. And in sparse factor modeling, post-processing is the key step that turns posterior draws into interpretable evidence about multimodality and non-uniqueness [Kaufmann and Pape, 2025]. Taken together, these contributions emphasize a practical criterion for Bayesian methods in econometrics and statistics: a model is not truly part of the methodological toolkit unless it can be fit with stable computation and yields interpretable posterior summaries at the scale where contemporary applications live.

With the special issue now complete, it is worth returning to the spirit of the original call. The collection does not simply demonstrate that Bayesian methods can be applied across diverse domains; rather, it shows how Bayesian modeling can generate new representations (of dependence, heterogeneity, and causal structure) and new inferential strategies (for truncation by death, semi-supervision, multi-dataset integration, and high-dimensional dynamics) that are difficult to obtain through routine extensions of standard tools.

Roberto Casarin
University Ca' Foscari of Venice
Email: r.casarin@unive.it

Luciana DallaValle
University of Plymouth, UK
Email: luciana.dallavalle@plymouth.ac.uk

Michele Guindani
University of California, Los Angeles, USA
Email: Mguindani@ucla.edu

Daniel Kowal
Rice University, USA
Email: daniel.kowal@rice.edu

Yasuhiro Omori
University of Tokyo, Japan
E-mail: omori@e.u-tokyo.ac.jp

References

Ovielt Baltodano Lopez, Federico Bassetti, Giulia Carallo, and Roberto Casarin. First-order integer-valued autoregressive processes with Generalized Katz innovations. *Econometrics and Statistics*, 2025. doi: 10.1016/j.ecosta.2025.03.003. URL https://doi.org/10.1016/j.ecosta.2025.03.003. In Press, Corrected Proof. Available online 12 March 2025.

Guillermo Granados-Garcia, Raquel Prado, and Hernando Ombao. Bayesian nonparametric multivariate mixture of autoregressive processes with application to brain signals. *Econometrics and Statistics*, 2024. doi: 10.1016/j.ecosta.2024.01.004. URL https://doi.org/10.1016/j.ecosta.2024.01.004. In Press, Corrected Proof. Available online 17 February 2024.

Giulio Grossi, Marco Mariani, Alessandra Mattei, and Fabrizia Mealli. Bayesian principal stratification with longitudinal data and truncation by death. *Econometrics and Statistics*, 2025. doi: 10.1016/j.ecosta.2025.01.003. URL https://doi.org/10.1016/j.ecosta.2025.01.003. In Press, Corrected Proof. Available online 6 February 2025.

Iván Gutiérrez, Danilo Alvares, and Luis Gutiérrez. A Bayesian flexible model for testing Granger causality. *Econometrics and Statistics*, 2024. doi: 10.1016/j.ecosta.2024.08.001. URL https://doi.org/10.1016/j.ecosta.2024.08.001. In Press, Corrected Proof. Available online 3 August 2024.

Sylvia Kaufmann and Markus Pape. Bayesian (non-)unique sparse factor modeling. *Econometrics and Statistics*, 2025. doi: 10.1016/j.ecosta.2025.04.001. URL https://doi.org/10.1016/j.ecosta.2025.04.001. In Press, Corrected Proof. Available online 23 April 2025.

Jinlu Liu, Sara Wade, and Natalia Bochkina. Shared differential clustering across single-cell RNA sequencing datasets with the hierarchical Dirichlet process. *Econometrics and Statistics*, 2024. doi: 10.1016/j.ecosta.2024.02.001. URL https://doi.org/10.1016/j.ecosta.2024.02.001. In Press, Corrected Proof. Available online 17 February 2024.

Gözde Sert, Abhishek Chakrabortty, and Anirban Bhattacharya. Bayesian semi-supervised inference via a debiased modeling approach. *Econometrics and Statistics*, 2025. doi: 10.1016/j.ecosta.2025.05.001. URL https://doi.org/10.1016/j.ecosta.2025.05.001. In Press, Corrected Proof. Available online 13 May 2025.

Seyma Vahap. Bayesian dynamic graphical models for high-dimensional vector autoregressions with time-varying parameters and volatility discounting. *Econometrics and Statistics*, 2025. doi: 10.1016/j.ecosta.2025.07.001. URL https://doi.org/10.1016/j.ecosta.2025.07.001. In Press, Corrected Proof. Available online 19 July 2025.