PSYCHOMETRIKA—VOL. 87, NO. 2, 373–375
JUNE 2022
https://doi.org/10.1007/s11336-022-09850-0

# GUEST EDITORS' INTRODUCTION TO THE SPECIAL ISSUE ON FORECASTING WITH INTENSIVE LONGITUDINAL DATA

PETER F. HALPIN

THE UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL

KATHLEEN GATES

UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL

SIWEI LIU

UNIVERSITY OF CALIFORNIA AT DAVIS

The collection of longitudinal data in the social and behavioral sciences has been revolutionized by the widespread availability of information technologies such as smartphones, wearable technology, social media, digital learning, online games, and the Internet more generally. We use the term intensive longitudinal data (ILD) inclusively, to refer to the types of data available from such sources. ILD can arise from a broad range of data collection methods and research designs, and typically result in multivariate observations collected from multiple respondents over a relatively large number of time points. Examples of ILD from this collection of papers include experience sampling of alcohol and substance use, daily diaries of emotional states, students' interactions with a web-based math tutoring app, and university attendance records.

ILD are typically collected on an ongoing basis over an extended duration of time, and in some applications data collection may continue indefinitely. This introduces the prospect of analyzing and acting upon ILD as they arrive, rather than waiting for data collection to be completed. This "online" approach to data analysis is common in domains such as engineering and machine learning, where it is usual for data to arrive on an ongoing basis. However, its potential application in the social and behavioral sciences remains largely unexplored. An initial step in this direction is to address the problem of forecasting with ILD, which is the focus of this special collection.

The statistical analysis of ILD has motivated the development of novel modeling approaches. Examples from previous literature that are considered in this collection include multilevel/random-effects extensions of vector auto-regression (VAR), dynamical structural equation modeling (SEM), and (non-) linear dynamical systems models. Hunter et al. (2022) provide an overview (and accompanying software) for filtering and forecasting techniques used in dynamical systems. Chow et al. (2022) consider how dynamical systems can be combined with control theory to "steer" a system towards a desired state, and consider the implications of this approach for personalized education. Lafit et al. (2021) provide data-driven insights into several factors that affect the predictive accuracy of multilevel VAR. This collection of papers also introduces some new modeling approaches. Li et al. (2022) introduce a multilevel zero-inflated Poisson (ZIP) model in which both the Poisson counts and the ZIP regimes have person-specific auto-regressive time dependency. Fisher et al. (2022) introduce an alternative to the random-effects approach, instead using regularization (adaptive LASSO) to extend VAR to multiple subjects while ensuring sparsity of the resulting solution. Additional approaches include a latent class extension of dynamic

Correspondence should be made to Peter F. Halpin, School of Education, The University of North Carolina at Chapel Hill, 100 E Cameron Ave, Chapel Hill, NC27599-3500, USA. Email: peter.halpin@unc.edu



© 2022 The Author(s) under exclusive licence to The Psychometric Society

https://doi.org/10.1007/s11336-022-09850-0 Published online by Cambridge University Press

SEM (Keleva et al. 2022) and machine learning extensions of mixed-effects models (Nestler & Hamburg, 2021)

One theme that arises from this collection is the importance of disentangling intra- and inter-individual sources of variation. While this general theme is certainly familiar, it leads to some interesting considerations in the context of forecasting and prediction as distinct from estimation and inference. Lafit et al. (2021) nicely characterize this tension in the context of multilevel VAR. More generally, while there is a consensus that models for ILD should accommodate individual-specific parameterizations (e.g., via random effects), there are open questions as to the conditions under which prediction will be more accurate using individual-specific versus aggregate model properties (e.g., the predicted random effects or the fixed effects). The comparison of the forecasting properties of aggregate and individual-specific model specifications, such as those provided by Nestler and Hamburg (2021) as well as Li et al. (2022), begins to address this question.

A second theme concerns how to explicitly link forecasting quality to model parameters. Often the forecasting distribution is not available analytically. In such cases it is difficult to understand how model parameters influence forecasts, or how changing a parameter (e.g., intervening on a system) affects the forecast distribution. Chow et al. (2022) discuss this problem from the perspective of control theory. Lafit et al. (2021) provide insights about how both model and data characteristics affect prediction with multilevel VAR, while Nestler and Hamburg (2021) discuss some analytical results in the context of mixed-effects models. In many applications, accurate forecasts may lead to an appropriate course of action even if we lack a precise understanding of how those actions affect a system's dynamics (e.g., in the case of alcohol and substance use). Nonetheless, learning about these effects through the model is certainly one way in which statistical research can inform substantive theories about ILD.

Other than the contribution by Hunter et al. (2022), the papers in this collection implemented "batch" estimation in the usual way (i.e., after the data were in). While online estimation is not required for forecasting, in its absence one must address the question of how estimation and forecasting can be integrated in applications that involve ongoing data collection. Perhaps the model is "pre-calibrated" using an initial batch of data and then used for forecasting without updating model parameters, which is an inefficient use of data. Alternatively, the model might be "naively" re-estimated after each data point is observed, which is an inefficient (and perhaps infeasible) use of computational resources. Integration of estimation and forecasting via online approaches whose properties are well understood remains a long-term goal for advancing the study of ILD.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Chow, S-M., Lee, J., Hofman, A. D., van der Maas, H. L. J., Pearl, D. K., & Molenaar, P. C. M. (2022). Control theory forecasts of optimal training dosage to facilitate children's arithmetic learning in a digital education application. *Psychometrika*. https://doi.org/10.1007/s11336-021-09829-3.

Fisher, Z. F., Kim, Y., Fredrickson, B., & Pipiras, V. (2022). Penalized estimation and forecasting of multiple subject intensive longitudinal data. *Psychometrika*. https://doi.org/10.1007/s11336-021-09825-7.

Hunter, M. D., Fatimah, H., & Bornovalova, M. A. (2022). Two filtering methods of forecasting linear and nonlinear dynamics of intensive longitudinal data. *Psychometrika*. https://doi.org/10.1007/s11336-021-09827-5.

Kelava, A., Kilian, P., Glaesser, J., Merk, S., & Brandt, H. (2022). Forecasting intra-individual changes of affective states taking into account interindividual differences using intensive longitudinal data from a university student drop out study in math. *Psychometrika* (current issue).

Lafit, G., Meers, K., & Ceulemans, E. (2021). A systematic study into the factors that affect the predictive accuracy of multilevel VAR(1) models. *Psychometrika*. https://doi.org/10.1007/s11336-021-09803-z.

Li, Y., Oravecz, Z. Zhou, S., Bodovski, Y., Barnett I. J., Chi, G., Zhou, Y., Friedman, N. P., Vrieze, S. I., & Chow, S-M. (2022). Bayesian forecasting with a regime-switching zero-inflated multilevel Poisson regression model: an

https://doi.org/10.1007/s11336-022-09850-0 Published online by Cambridge University Press

application to adolescent alcohol use with spatial covariates. *Psychometrika*. https://doi.org/10.1007/s11336-021-09831-9.

Nestler, S., & Humberg, S. (2021). A Lasso and a regression tree mixed-effect model with random effects for the level, the residual variance, and the autocorrelation. *Psychometrika*. https://doi.org/10.1007/s11336-021-09787-w.

*Accepted: 29 JAN 2022*
*Published Online Date: 1 MAR 2022*

https://doi.org/10.1007/s11336-022-09850-0 Published online by Cambridge University Press