De : Marc A Scott <marc.scott@nyu.edu>
Envoyé : vendredi 18 janvier 2019 04:52
À : Swisspanel <swisspanel@fors.unil.ch>
Cc : Jacques-Antoine Gauthier <jacques-antoine.gauthier@unil.ch>; Kaushik Mohan <km4326@nyu.edu>
Objet : Abstract for 10th International Conference of Panel Data Users in Switzerland

To the organizing committee:

I would like to submit the following abstract for a presentation at the 10th International Conference of Panel Data Users at the University of Lausanne on 5-7 June, 2019. Thank you for your consideration.

Names and affiliations of all authors:
Marc Scott (New York University)
Kaushik Mohan (New York University)
Jacques-Antoine Gauthier (Université de Lausanne)

The names and e-mail addresses of the attending authors:
Marc Scott (marc.scott@nyu.edu)
Jacques-Antoine Gauthier (Jacques-Antoine.Gauthier@unil.ch)

The session to which the abstract belongs:
Longitudinal methods (Gauthier J.-A. and Le Goff, J.-M., organizers)

Title:
Hybrid clustering models for lifecourse studies with application to the Swiss Household Panel

Abstract:
Methods and models for longitudinal data with single, continuous outcomes are fairly well established, but less so when the outcomes of interest are categorical and/or multidimensional, as they often are in lifecourse studies. Exceptions include when the data are binary or perhaps ordinal, for which models analogous to those in the continuous case have been developed. When faced with a larger number of states, or when the categorical outcomes are multidimensional, two types of methods are typically employed: clustering algorithms relying on a distance metric based on Optimal Matching; or parametric models of duration in a competing risks framework. The multidimensionality is addressed using "multi-channel" approaches or by extending the state space via a cross product. A longstanding debate as to which approach is appropriate derives from competing goals: clustering sorts the population into typologies; duration models are organized around individuals and their time-dependent choice sets. Implicit in this debate is the necessity, or not, of utilizing a data generating process (DGP) and how one might incorporate a DGP into the life course perspective, in which the unit of analysis is represented as a chronological sequence of statuses.

A DGP offers some significant advantages over non-model approaches: one can generate implied realizations from the DGP, build confidence bounds from them, and compare their properties to those of the given data, akin to posterior predictive model checking. This introduces a new set of tools for establishing goodness of fit. Moreover, one can

systematically add components to the DGP to reflect various aspects of the real-world process, e.g., the timing and conditionality of events. However, when the goal is discovery of groups of similar profiles, use of optimal matching techniques (OMA) has been extremely successful. Viewed from a distance, OMA actually corresponds roughly to a "perturbed medoid" model, with the perturbation mechanism based on token co-occurrence properties extant in the data. The perturbations in this model are harder to interpret sociologically, but the model is quite effective at establishing groups with similar trajectories.

We begin to unify the approaches by examining increasingly complex (and thus realistic) latent class models for sequences. These are a hybrid class of growth mixture models with an error process that captures the tendency to remain in a given state (e.g., cohabitation or employment) in a natural manner. We incorporate an individual, competing-risks duration model within a population-level model-based clustering framework. We examine this model class in the context of the Swiss Household Panel (SHP), exploring the challenge of multi-channel sequences, building on the work of Gauthier et al. (2010). We show that building a combined individual and group level model for the complex dependence structure provides new insights into the social process, as well as highlights the potential benefits of an integrative, model-based technique.

Sincerely,
Marc Scott