COMPLEX TRAJECTORIES

Group Based Trajectory Modelling

May 2023

Co-funded by the Erasmus+ Programme of the European Union

Author:

Janine Jongbloed

The following documents are embedded in a whole MOOC course on Longitudinal analysis techniques developed in the framework of the same Complex Trajectories project. Those interested in doing the MOOC should consider the following procedure:

1. Access the AULAbERTA space through the link: https://aulaberta.uab.pt/

2. Register in the platform a. Select one of the available languages (Portuguese or English) b. Follow the instructions given in the platform

3. After creating the account, they should access the MOOC Longitudinal Analysis through the link: https://aulaberta.uab.pt/course/view.php?id=94

Co-funded by the Erasmus+ Programme of the European Union

Group-based trajectory modeling

Janine Jongbloed, IREDU, Université de Bourgogne

Co-funded by the Erasmus+ Programme of the European Union

Hello and welcome to this unit focused on group-based trajectory modeling.

Developmental trajectory groups

- **Introduction/brief history**
- **Methodological approach**
  - Assumptions and groups as 'useful fiction'
  - Model selection and fit
  - Linking trajectories to covariates
  - Dual vs. multi-trajectory analysis
- Applications in educational research

Co-funded by the Erasmus+ Programme of the European Union

In this first video, I'll give a very brief introduction, followed by a step-by-step explanation of the methodological approach. In the next video, I'll share some interesting applications of this approach in educational research.

Introduction

- Originally created for use in **criminology** (Nagin and Land, 1993)
  - developmental origins of crime and delinquency and the "criminal career model"
  - "person-based" life-course perspective on physical aggression
- Charts longitudinal "**developmental trajectories**"
  - describes and helps explain changes over a relatively long period of time
  - summarizes individual differences in the developmental progression of a variable
- Categorical trajectory groupings to **reduce complexity**
  - we assume there are meaningful (homogeneous) subgroups that follow distinctive trajectories
  - "heuristic summary" for approximating differences in developmental trajectories

Co-funded by the Erasmus+ Programme of the European Union

Nagin, Daniel. 2005. *Group-Based Modeling of Development*. Harvard University Press. doi:10.4159/9780674041318.

Group-based trajectory modeling was originally used in criminology in the early 1990s by Daniel Nagin to describe criminal career models and a give a person-based longitudinal perspective on criminality. To do so, Daniel Nagin charted developmental trajectories of individual behaviour to see the progression of a variable over time, and then grouped these individual trajectories into groups of individual with similar developmental patterns to find sub-groups that follow distinctive trajectories over time.

Methodological approach

Co-funded by the Erasmus+ Programme of the European Union

Next, we'll explore the data, assumptions, and steps involved in this approach.

Modeling assumptions

- Finite mixture modeling using maximum likelihood estimation (Nagin, 2005)
  - e.g., Poisson, logit and tobit (censored normal distribution) regression
  - "semi-parametric" approach using finite number of discrete groups to approximate a continuous distribution
- Charts longitudinal developmental trajectories
  - estimates a set of parameters to define the shapes of the trajectories (polynomial function of time) and the probability of trajectory group membership
- Relies on the "conditional independence assumption"
  - for individuals in the same group, outcomes over time "are assumed not to be serially correlated in the sense that individual-level deviations from the group trend are uncorrelated" (to reduce model complexity)

Nagin, Daniel. 2005. *Group-Based Modeling of Development*. Harvard University Press. doi:10.4159/9780674041318.

Co-funded by the Erasmus+ Programme of the European Union

Group-based trajectory modeling uses a finite mixture model with maximum likelihood estimation. Our variables could be either continuous or binary, and ordinal scale variables, such as those from psychological scales, can also be used. Our aim is to model the shape of the trajectories over time, so we typically need at least 3 points in time or more for each individual.

Modeling assumptions

- We are interested in the distribution of outcomes conditional on age or time (p. 28)

$$P(Y_i|\text{Age}_i) = \sum_{j=1}^{J} \pi^j \times P(Y_i|\text{Age}_i, j; \beta^j), \qquad (1)$$

$$P(Y_i|\text{Age}_i, j; \beta^j) = \prod_{t=i}^{T} p(y_{it}|\text{age}_{it}, j; \beta^j), \qquad (2)$$

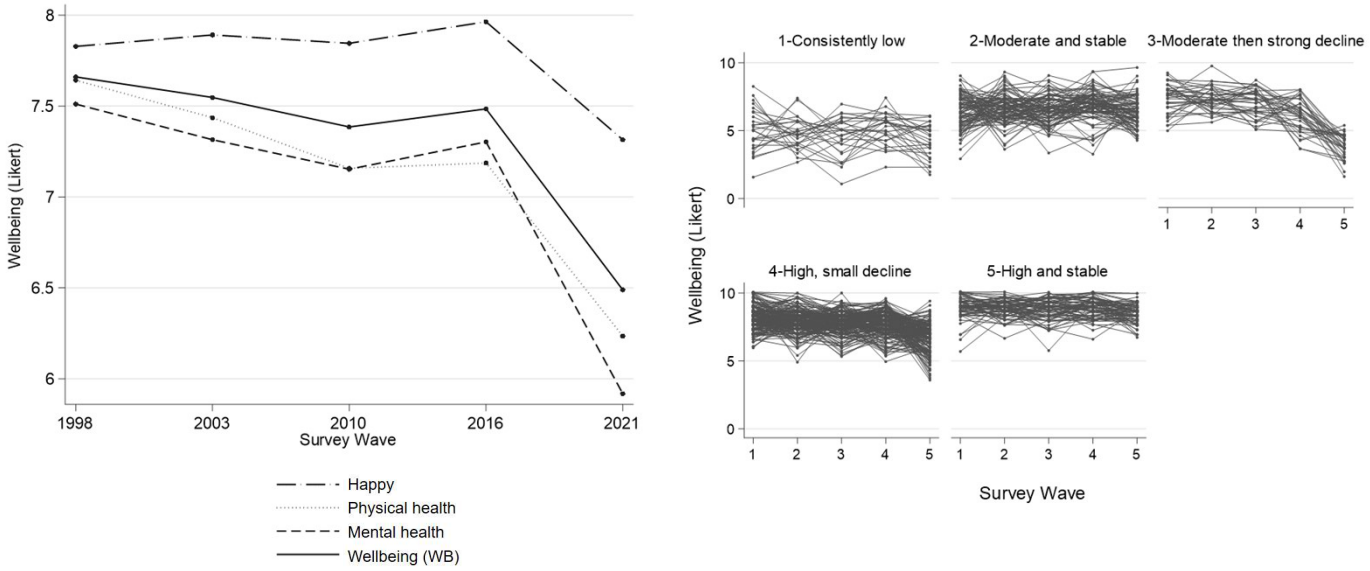
- Different from conventional growth curve model
  - growth curve modeling describes population mean with individual trajectories varying around this mean (multivariate normal distribution)
  - GBTM takes a "multinomial" approach to growth processes

Nagin, Daniel. 2005. *Group-Based Modeling of Development*. Harvard University Press. doi:10.4159/9780674041318.

Co-funded by the Erasmus+ Programme of the European Union

You can find all the details about the statistical assumptions made in Daniel Nagin's 2005 book *Group-Based Modeling of Development.* In essence, this approach is a multinomial approach to growth curve modeling, that allows individual trajectories to vary across multiple group averages, rather than only one.

Complex Trajectories



Jongbloed, J. and Andres, L. (2023) Charting well-being over adulthood into pandemic times: a longitudinal perspective, *Longitudinal and Life Course Studies*, DOI: 10.1332/175795921X16715373405417

Co-funded by the Erasmus+ Programme of the European Union

7

Here, for example, we see well-being scores on a composite scale shown over time. On the left, we see the average scores for the whole sample across time, while on the right, we see individual scores plotted across five trajectory groups, as defined by a group-based trajectory modeling approach. We see that each of these groups shows a different pattern, and this pattern is different among groups, and different from the overall sample averages as well. Notably, the dip in well-being that we see in the overall average scores does not appear in all of the trajectory groups. Some groups of individuals do not shown this « sample average » pattern.

'How-to'

1. Create a hypothesis of a plausible number of groups based on theory/literature
2. Refine the model from step 1 to determine:
   a) the optimal number of groups, typically testing K=1–7 groups
   b) the optimal shape of the trajectories, typically testing linear, quadratic, and cubic functions of each trajectory
3. Assess model fit using Bayesian information criteria (BIC) values, average posterior probability of assignments (APPA), and odds of correct classification (OCC)
4. Investigate graphical presentations and assess for substantive interpretation

Lennon, Hannah, Scott Kelly, Matthew Sperrin, Iain Buchan, Amanda J. Cross, Michael Leitzmann, Michael B. Cook, and Andrew G. Renehan. 2018. "Framework to Construct and Interpret Latent Class Trajectory Modelling." *BMJ Open* 8 (7). doi:10.1136/bmjopen-2017-020683.

Co-funded by the Erasmus+ Programme of the European Union

To conduct group-based trajectory modeling, we first need to hypothesize a plausible number of groups based on the shapes of trajectories that are suggested by theory or in the literature. Next, we refine our model by testing different numbers of groups and different shapes of the trajectories. We have a number of statistical measures that we can use to assess the fit of our model. Finally, we can look at the graphs of our models to see if our results are consistent with our hypotheses.

Optimal # of groups

- Bayesian information criteria (BIC) value
  - test nested models to find one with the maximum (i.e., least negative) BIC score that contains no more groups than is necessary to capture the distinct features of the data (**parsimony**)
  - then test fit of model using (=exp(BIC*i* - BIC*j*)) where model *j* has one more group than model *i*

Value	Evidence for more complex model
0-2	Weak
2-6	Moderate
6-10	Strong
10+	Very strong

Andruff, Heather, Natasha Carraro, Amanda Thompson, Patrick Gaudreau, and Benoît Louvet. 2009. "Latent Class Growth Modelling: A Tutorial." *Tutorials in Quantitative Methods for Psychology* 5 (1): 11–24. doi:10.20982/tqmp.05.1.p011.

Co-funded by the Erasmus+ Programme of the European Union

The first measure that we use to test the optimal number of groups, is the bayesian information criteria (or BIC) value. We test nested models to find one with a strong value, but that contains no more groups than necessary to capture the distinct features in the data. We use this measure in collaboration with our next measures in an iterative approach.

Group assignment

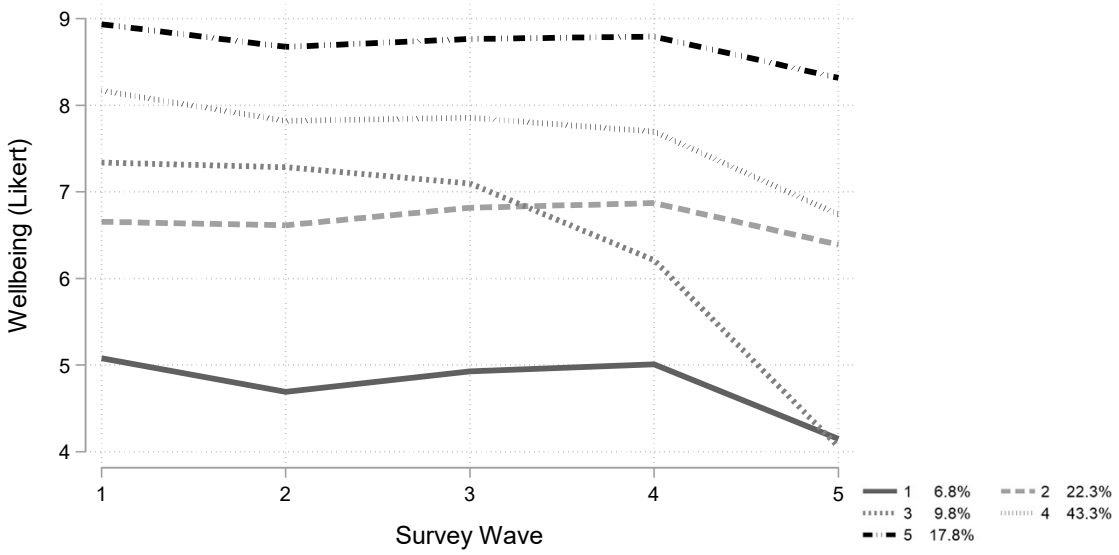
- Average posterior probabilities (APP) of group assignment
  - an approximation of the internal reliability for each trajectory group
  - averages the posterior probabilities of individuals having been assigned group membership to a trajectory using a maximum probability assignment rule
  - Nagin (2005) recommends that each group have an APP of at least **0.7**
- The odds of correct assignment should exceed **5**
- Ideally, each trajectory group should also have an approximate group membership probability of at least **5** %

Andruff, Heather, Natasha Carraro, Amanda Thompson, Patrick Gaudreau, and Benoît Louvet. 2009. "Latent Class Growth Modelling: A Tutorial." *Tutorials in Quantitative Methods for Psychology* 5 (1): 11–24. doi:10.20982/tqmp.05.1.p011.

Co-funded by the Erasmus+ Programme of the European Union

We use the measure of average posterior probabilities of group assignment to test the internal reliability of each group, using a cut-off value of 0.7, and an estimate of the odds of correct assignment with a value of at least 5. We also aim to have at least five percent of the sample in each of the groups, to avoid having very small groups of only a few individuals. Using these four criteria, we compare different group solutions to arrive at the best one, which also needs to be theoretically interpretable.

Complex Trajectories



Jongbloed, J. and Andres, L. (2023) Charting well-being over adulthood into pandemic times: a longitudinal perspective, *Longitudinal and Life Course Studies*, DOI: 10.1332/175795921X16715373405417

Co-funded by the Erasmus+ Programme of the European Union

11

For example, when we look again at well-being scores over time, the best-fitting trajectory group solution is five groups. Group 1 shows consistently low well-being scores across time, and includes 7 percent of the sample, Group 2 shows moderate and stable scores for 22 percent of the sample, Group 3 shows moderate well-being scores followed by a strong decline for 10 percent of the sample, Group 4 shows high scores with a small decline for more than 40 of the sample, and Group 5 shows high and relatively stable well-being scores across time for almost 20 percent of the sample. In this study, we used this evidence of a diversity in well-being trajectories to challenge the idea of a generalized midlife decline in well-being.

Including covariates

- Coefficients are interpreted like those in conventional regression, as either:
  a) the association of the predictor variables with trajectory group membership probability, or
  b) a measure of the predicted change in the response variable *y* in a given year and for a given trajectory group for a given change in an explanatory variable
- Time-invariant covariates–
  - individual-level characteristics
  - predict probability of trajectory **group membership** using multinomial logit model
- Time-varying covariates–
  - events that occur during a trajectory
  - predict if they are related to the **shape of the trajectory**

Nagin, Daniel. 2005. *Group-Based Modeling of Development*. Harvard University Press. doi:10.4159/9780674041318.

Co-funded by the Erasmus+ Programme of the European Union

A further interesting application of group-based trajectory modeling is to include covariates to predict trajectory group membership probabilities or to predict the effect of group membership on a given outcome. We can use both time-invariant or time-varying variables, depending on our research questions.

Joint/dual trajectory analysis

- We can also investigate the probability of following a specific trajectory for outcome *B* given the individual is following a specified trajectory for outcome *A*
- To do so, we compute:
  1. trajectory groups for both measurement series;
  2. the probability of membership in each such trajectory group; and
  3. **probabilities** linking membership in trajectory groups across behaviors

Nagin, Daniel S., Bobby L. Jones, Valéria Lima Passos, and Richard E. Tremblay. 2018. "Group-Based Multi-Trajectory Modeling." *Statistical Methods in Medical Research* 27 (7): 2015–2023. doi:10.1177/09

Co-funded by the Erasmus+ Programme of the European Union

We can also extend group-based trajectory modeling to look at multiple developmental trajectories (for example, two different psychological scales or behaviours), and to see whether a shape of trajectory on one scale is linked to a particular shape of trajectory on a different scale.

Multi-trajectory analysis

- Another option is to define trajectory groups in terms of trajectories for multiple (not just one) outcome
- To do so, we estimate:
  1. trajectory models with varying number of groups for each of the outcomes separately in order to clarify the types of distinctive trajectories;
  2. a multi-trajectory model having *at least* the number of groups found in step 1; and
  3. **model fit** using conventional criteria

Nagin, Daniel S., Bobby L. Jones, Valéria Lima Passos, and Richard E. Tremblay. 2018. "Group-Based Multi-Trajectory Modeling." *Statistical Methods in Medical Research* 27 (7): 2015–2023. doi:10.1177/09

Co-funded by the Erasmus+ Programme of the European Union

This extension can also be applied in a slightly different manner to more than two groups, where we can define trajectories based on multiple variables or scales and combine them to find an overall number of groups. If you're interested in these advanced applications, I encourage you to refer to the articles listed here and on the previous slides.

Thank you!

Please refer to the supporting materials on the MOOC website.

15

Thank you for your attention, and see you in the next video, where we will examine several applications of these techniques in educational research.

Group-based trajectory modeling

Janine Jongbloed, IREDU, Université de Bourgogne

Co-funded by the Erasmus+ Programme of the European Union

Hello again, and welcome back to unit 4 focused on group-based trajectory modeling.

Developmental trajectory groups

- Introduction/brief history
- Methodological approach
  - Assumptions and groups as 'useful fiction'
  - Model selection and fit
  - Linking trajectories to covariates
  - Dual vs. multi-trajectory analysis
- **Applications in educational research**

Co-funded by the Erasmus+ Programme of the European Union

In our last video, I presented an overview of the logic and steps of group-based trajectory modeling. Next, we'll look at examples from several published research articles in the field of educational studies.

Applications in educational research

Co-funded by the Erasmus+ Programme of the European Union

Surprisingly, there are relatively few studies in education that apply this technique, although the examples from criminology and psychology are numerous. We will focus on two particularly interesting articles.

Persistence patterns over 3 years

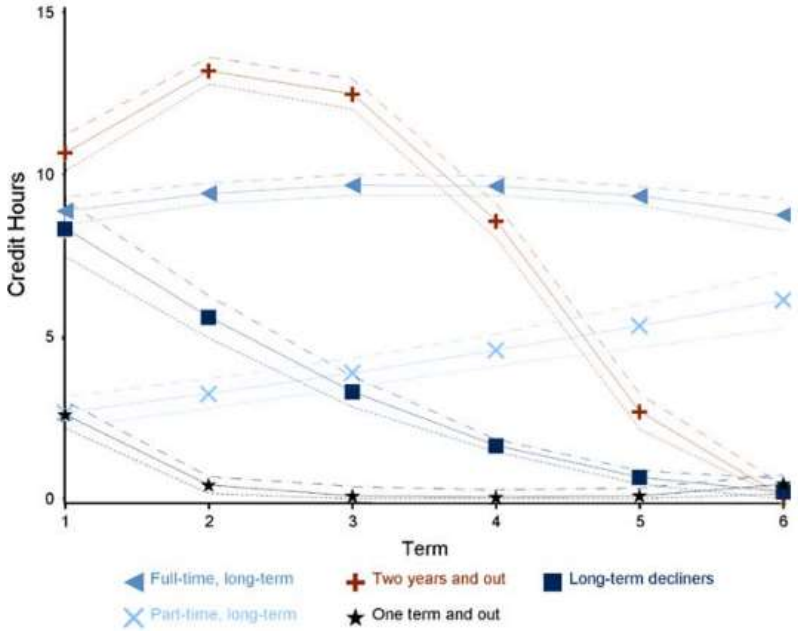
- Much variation in community college students' postsecondary pathways
  - some pathways eventually lead to academic success but "difficult to classify as successes in terms of conventional accountability measures" (p. 318)
  - need better understanding of **persistence/enrollment patterns** → taxonomy
- Subjective categorizations of enrollment patterns in the literature
  - diverse educational patterns (Adelman 2005; Bailey et al. 2004, Goldrick-Rab 2006; McCormick 2003)
  - multiple distinct patterns will provide better model than variability around a central tendency
- Credit hours completed = successful persistence

Nathan Marti, C. 2008. "Latent Postsecondary Persistence Pathways: Educational Pathways in American Two-Year Colleges." *Research in Higher Education* 49 (4): 317–336. doi:10.1007/s11162-007-9083-2.

Co-funded by the Erasmus+ Programme of the European Union

Our first example, a study by Nathan Marti published in 2008, looked at pathways through post-secondary education over three years. He decided to model the number of credit hours completed over time to capture different types of persistence in enrollment patterns.

Complex Trajectories



Nathan Marti, C. 2008. "Latent Postsecondary Persistence Pathways: Educational Pathways in American Two-Year Colleges." *Research in Higher Education* 49 (4): 317–336. doi:10.1007/s11162-007-9083-2.

Co-funded by the Erasmus+ Programme of the European Union

5

He found five groups of persistence patterns, including those who enrolled long-term either full-time or part-time, those who dropped out after one or two terms, and those whose participation slowly declined over the three-year period.

Persistence patterns over 3 years

- Five distinct patterns emerged in three independent samples
  - consistencies with theoretical groupings (Adelman 2005; Bailey et al. 2004, Goldrick-Rab 2006; McCormick 2003)
- Significant differences between trajectory groups in their levels of **engagement**
- Students' use of institutions "rarely fits a standard timeline and enrollment pattern" (p. 327)
- Limitations: 3 years; single institutions
  - need multi-institutional data to know whether a students' departure from an institution represents a successful or unsuccessful outcome

Co-funded by the Erasmus+ Programme of the European Union

Nathan Marti, C. 2008. "Latent Postsecondary Persistence Pathways: Educational Pathways in American Two-Year Colleges." *Research in Higher Education* 49 (4): 317–336. doi:10.1007/s11162-007-9083-2.

These five groups were also supported theoretically in the literature, and showed that real student participation often deviates from the institutional timelines and expectations. To read more about this study, you can refer to the article listed here.

Persistence patterns over 10 years

- National Longitudinal Survey of Youth 1997, 10 years
- Group-based developmental trajectories of students' completed course credits over time
  - "the traditional pathway… characterizes the experience of a minority of students" (Milesi, 2010)
- How student characteristics predict trajectory group membership
  - students' socioeconomic and prior academic backgrounds
  - transitions to other adult social roles (spouse, parent, and worker)
- Postsecondary educational participation from a life course perspective
  - "a process characterized by steady and patterned changes in enrollment status and credit accumulation that can unfold over a relatively long period of time" (p. 244)

Co-funded by the Erasmus+ Programme of the European Union

Denice, Patrick. 2019. "Trajectories through Postsecondary Education and Students' Life Course Transitions." *Social Science Research* 80 (January). Elsevier: 243–260.

A second more recent study by Patrick Denice in 2019 also looked at trajectories through post-secondary education, this time over a period of ten years. He also examines how student characteristics predict their participation patterns over early adulthood.

Persistence patterns over 10 years

- Credits completed per year = successful persistence
- Includes timing, duration, and sequencing of students' progress:
  i. Whether a student is enrolled,
  ii. How long they waited after high school to enroll,
  iii. Whether and for how long they persist in PSE, and
  iv. Whether they are enrolled part-time or full-time
- Explanatory variables (covariates)
  - to predict which trajectories an individual with certain characteristics is most likely to follow

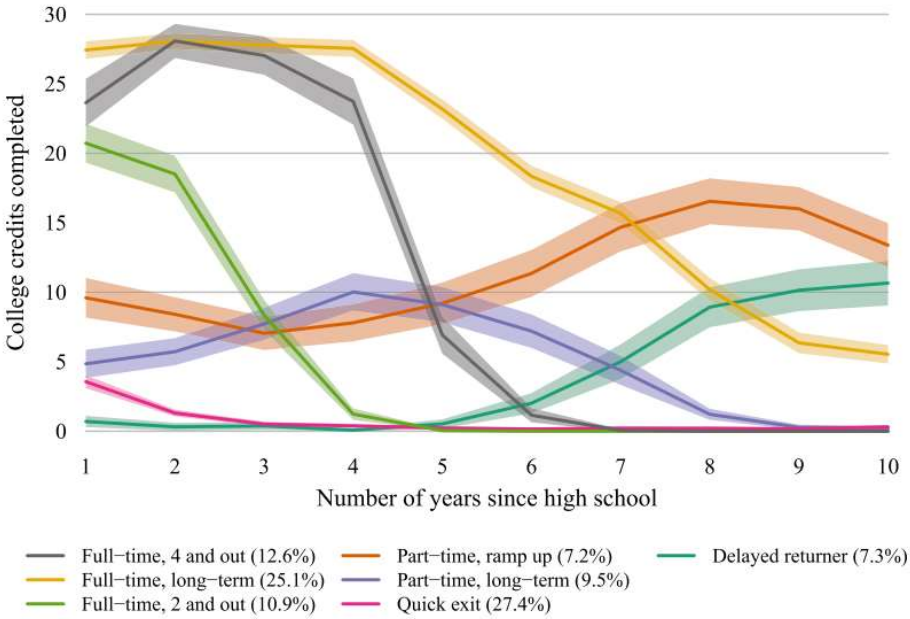
a) time-invariant: gender, race/ethnicity, socioeconomic status, and prior academic preparation
b) time-varying: marriage or cohabiting partnership, parenthood, and work (lagged by 1 year)

Co-funded by the Erasmus+ Programme of the European Union

Denice, Patrick. 2019. "Trajectories through Postsecondary Education and Students' Life Course Transitions." *Social Science Research* 80 (January). Elsevier: 243–260. doi:10.1016/j.ssresearch.2019.02.005.

In this study, completed course credits are once again used as the measure of persistence. This study aimed to capture the diversity in post-secondary attendance patterns over time incorporating a reflection on whether a student is enrolled, how long they waited after high school to enrol, how long they persist in post-secondary education, and whether they participate full-time or part-time across the period in question.
The trajectory groups were then linked to individual characteristics that are both constant, such as prior socio-economic status, or varying, such as family formation.

Complex Trajectories



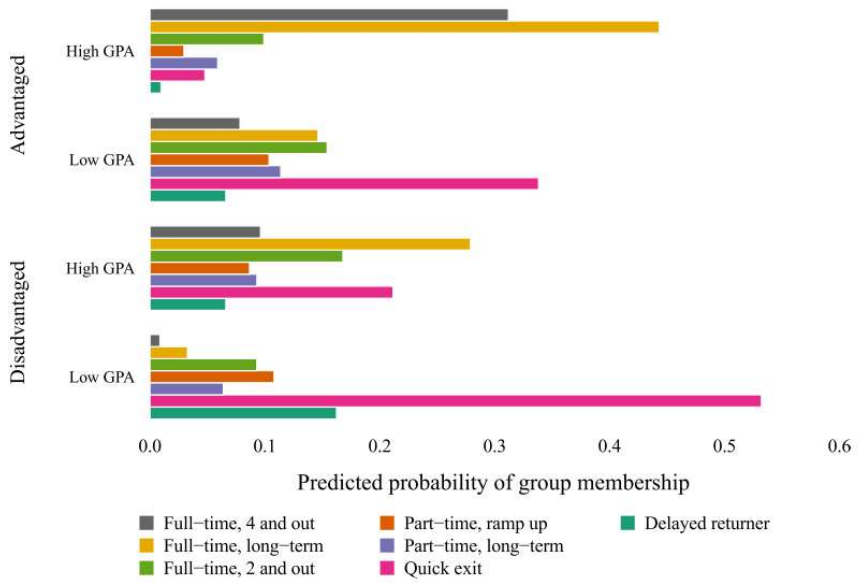
Denice, Patrick. 2019. "Trajectories through Postsecondary Education and Students' Life Course Transitions." *Social Science Research* 80 (January). Elsevier: 243–260. doi:10.1016/j.ssresearch.2019.02.005.

Co-funded by the Erasmus+ Programme of the European Union

9

Here we see the seven groups that Patrick Denice describes, including a traditional « four-year » group, those who participate over the long term either full- or part-time, those who delay their participation, those who stop either quickly or after two years, and those who increase their participation later in young adulthood.

Complex Trajectories



Denice, Patrick. 2019. "Trajectories through Postsecondary Education and Students' Life Course Transitions." *Social Science Research* 80 (January). Elsevier: 243–260. doi:10.1016/j.ssresearch.2019.02.005.

Co-funded by the Erasmus+ Programme of the European Union

10

He finds different combinations of variables can make students more likely to follow a particular type of participation pattern. When we compare those from socio-economically advantaged and disadvantaged backgrounds with high or low prior GPA scores, we see that those with a high GPA from a disadvantaged background are morel likely to have a « quick exit » pattern than those from an advantaged background. Likewise, those with a high GPA from a disadvantaged background are less likely to have a « long-term » participation pattern than those from an advantaged background.

Persistence patterns over 10 years

1. Patterns in trajectories
   - "Students follow multiple distinct pathways that rarely conform to standard timelines and that do not merely represent individual deviations from a single mean trend" (p. 255)
2. Individual characteristics
   - Those from disadvantaged socioeconomic backgrounds more likely to follow part-time enrollment, delay, and short patterns
3. Other time-varying adult roles
   - More employment hours linked to fewer completed credits

Co-funded by the Erasmus+ Programme of the European Union

Denice, Patrick. 2019. "Trajectories through Postsecondary Education and Students' Life Course Transitions." *Social Science Research* 80 (January). Elsevier: 243–260. doi:10.1016/j.ssresearch.2019.02.005.

These examples shows the utility of this approach: we can not only group patterns in trajectories that reflect a diversity of participation patterns, but we can also link individual characteristics to these patterns both over the whole trajectory and dynamically across specific points in time.

How can we build on these?

1. Country comparisons of trajectory groups
2. Visualize in more detail the most complex trajectories
   - for example, using sequence analysis in a second step
   - this was done recently by Virtanen et al. (2021)
     - the authors looked at months in employment over the course of each year for 7 years to determine employment trajectories
     - then they examined employment activities over this same time period for those in the "weak employment" group using sequence analysis (categories: student, employed, unemployed, other…)

Virtanen, Pekka, Arja Jolkkonen, Pertti Koistinen, Arja Kurvinen, Liudmila Lipiäinen, and Tapio Nummi. 2021. "Are the Early Leavers the Lucky Ones?" *Nordic Journal of Working Life Studies* 11 (1): 23–43. doi:10.18291/njwls.122199.

Co-funded by the Erasmus+ Programme of the European Union

These useful examples applied to national data allow us to imagine other possibilities for further research: Can we compare trajectory groupings across countries or educational systems? Also, can we combine this approach with sequence analysis, which we covered in the last unit, to uncover sub-groupings within larger participation patterns? A recent article did so on the topic of employment trajectories, and this could easily be adapted to educational research, as well.

Thank you!

Please refer to the supporting materials on the MOOC website.

13

Please refer to the articles cited throughout the presentation and the supporting MOOC materials for more details on each of the specific methodological approaches. Thank you for your attention!