Assessing the Degree of Linked Life Course Sequences of Family Dyads and Triads

Tim F. Liao, University of Illinois

Abstract

Dyadic/triadic life course sequences can be more associated within the dyad/triad than between them and other sequences, theoretically known as the life course principle of linked lives. In this paper, I propose a method for measuring and assessing the degree of linked life course trajectories in sequence data. Specifically, the distance between the members in an observed dyad/triad is compared against a set of randomly generated dyads/triads by assuming a particular random generation mechanisms. The proportion of such distance differences provides the degree of linked life courses in such dyadic/triadic sequence data. I present a simulation study and two empirical examples—an analysis of dyadic family formation using the Longitudinal Study of Generations data and an analysis of triadic employment history using the Panel Study of Income Dynamics data, to assess the performance of the method. The simulation study and the two applications demonstrate the usefulness of the proposed method.

Acknowledgements: The author wishes to acknowledge the benefits of the Discovery Project funded by the Australian Research Council and awarded to Irma Mooi-Reci, Mark Wooden, and Tim Liao. Working on the project initiated the research reported in this paper. The author would also like to thank Anette Fasang and Marcel Raab, who shared the data used in their 2014 *Demography* publication.

Introduction

This paper concerns the analysis of linked lives, especially how dyadic and triadic life course sequences are associated between family members of the same or different generations. The objective is to assess such an association as a measure of the degree of linked life courses between members of family dyads and triads.

Since its introduction from biology by Abbott and Forrest over three decades ago (1986), sequence analysis has been widely applied in the social sciences by the second-wave sequence analysis researchers (e.g., Aisenbrey and Fasang 2017; Fasang and Raab 2014). It has also witnessed rapid and continuing methodological advances in social sequence analysis (Barban et al 2017; Blanchard and Bühlmann 2014; Cornwell 2015; Fasang and Liao 2014; Studer 2013; Raab et al 2014; Studer, Struffolino, and Fasang 2018). In this paper, I follow up on this exciting research tradition of developing and applying sequence analysis.

There are three main approaches to analyzing dyadic sequence data. First, dyadic or triadic sequence data can be analyzed with multichannel sequence analysis, introduced by Gauthier et al. (2010). Fasang and Raab (2014), among others, provided a good example of such an application to parent-child family formation sequence data. Second, dyadic sequence data can be formed into the so-called grid-sequences, based on the state space grid method, and this approach is called grid-sequence analysis (Brinberg et al. 2016). Typically, cluster analysis follows multichannel sequence analysis or grid-sequence analysis, and the clusters generated often become the dependent variable in a subsequent substantive analysis. Third and finally, Liefbroer and Elzinga (2012) proposed a sub-sequence based approach to analyzing dyadic sequence data by focusing on the similarities of such subsequences based on optimal matching. There at least three differences between the proposed method in this paper and the three approaches above. While the current method also examines similarities of dyadic or triadic members, it is a relative kind of similarity in that the method compares the intra-dyad or intra-triad distance against simulated intra-dyad or intra-triad distances. This first difference implies a second difference: The proposed method provides an informative measure of the degree of linked life course sequences with a value for each and every dyad or triad. Finally, the proposed method has the flexibility of extending to the analysis of tetrads, pentads, and hexads, and to even higher dimensional polyads although such an extension is beyond the scope of the current paper.

The paper proceeds as follows. I first present a brief review of the linked lives concept, one of the five key life course principles. Next, I introduce the new method for measuring and analyzing the degree of linked life courses, for members of different (or same) generations. The method allows us to capture the degree of linked lives between members of (parent-child or sibling-sibling) dyads and members of (parents-child or siblings) triads. It also allows us to define a measured termed “significantly linked lives” or observed dyadic/triadic sequences resemble one another more than randomly generated sequences at least 95% times. I then introduce a simulation study of the degree of linked lives measure, before presenting two empirical examples, one analyzing the Longitudinal Study of Generations (LSOG) Survey family dyadic data and the other, the US Panel Study of Income Dynamics (PSID) family triadic data. Finally, I draw offer some reflections on possible future directions related to the linked life courses measure proposed here.

The Concept of Linked Lives

Among the five general life course principles of life-span development, agency, time and space, timing, and linked lives (Elder et al. 2003), the principle of linked lives is the only one that directly relates family members from different (or same) generations. As such, the concept of linked lives emphasizes the generational dimension of time in that one individual's life can be and most often is embedded within the lives of their family members, including those from other generations (Elder 1995; Macmillan and Copher 2005). Because individuals' lives are interdependent on others' lives in a family, their life course trajectories may in a way reflect the effects other members' life courses (Bengtson, Elder, and Putney, 2005). Thus, members' lives from different generations are uniquely connected across their own generation's life courses.

As reviewed and discussed by Gilligan et al. (2018), such interdependent, linked lives can create serious consequences such as cumulative inequality over generations. Parental job losses, for example, can have immediate outcome on another family member such as child's behavioral outcome. To study intergenerational dyads, the Family Transitions Project (formerly known as the Iowa Youth and Families Project) demonstrated how the Farm Crisis of the 1980s impacted parents' stress levels as well as the quality of family relationships between parents and children (Conger and Elder 1994). For life course research, the issue of linked lives is really about how one' life course is linked or associated with another's. The measurement of the degree of linked lives as represented in life course sequences such as these across generations is our focal point and our topic in the later sections.

Measuring Linked Life Courses

The principle of Linked lives is one of the most important concepts in life course research. However, to this day, there has not been a formal way for assessing or measuring the concept at the individual level (i.e., for every single dyad or triad). The proposed measure below is based on the principle of comparing observed and simulated dyadic and triadic life course sequences. Comparing observed and simulated data provides an effective statistical analysis for researchers in many disciplines (for application examples, see Amory et al. 2015; Furman et al. 2018).

The only statistical attempt related to the concept of linked lives in my best knowledge is an R package by Nightingale (2016) that computes how clusters such as members of households resemble one another as compared against randomly generated data. There are two limitations of this approach for analyzing life course sequences: First, the program produces a single statistic for the sample while ideally, a measure recording how linked the members' lives are in each cluster is desirable. Second and more important is the method for computing differences between observed and randomly generated data. It is based on differences only in the number of state changes whereas in sequence analysis a variety of dissimilarity or distance measures have been developed for analyzing life course sequences (Studner and Ritschard 2016).

The Procedure

The method described below overcomes both inadequacies. First, it evaluates the life course sequences in every linked dyad or triad against randomly generated dyad or triad based on a random generation assumption or mechanism (to be discussed later). Second, it computes distances using a particular distance measured sensitive to timing, duration, or order (Studer and Ritschard 2016). The method follows the principle of the permutation test or more broadly that of the randomization test as a nonparametric statistical method, as described in Liao (2002/2011). The method for analyzing both dyadic (e.g., parent and child or sibling-sibling) and triadic (e.g., two parents and a child or three siblings) sequence data follows the steps below:

1. a) Life course sequence data are collected into two (for dyadic data) or three (for triadic data) sets, each of which has N number of sequences, such as S_{1i}, S_{2i} or S_{1i}, S_{2i}, S_{3i}, for i = 1 to N;

   b) compute distances between the two or three sequences in observed sequences, using a user-defined dissimilarity measure, such as, for dyads:

   $$D_i = d(S_{1i}, S_{2i}) \qquad (1)$$

   where $d(\cdot)$ is a user-chosen distance function.

2. The generation of randomly chosen dyadic or triadic sequences:

   a) A randomly generated dyad or triad such as S_{1t}, S_{2t} or S_{1t}, S_{2t}, S_{3t}, for the tth simulation with length s by assuming a particular random generation mechanism is produced;

   b) compute distances between the two or three sequences in the randomly generated sequences in the same dyad or triad, using a user-defined dissimilarity measure, such as, for dyads:

   $$D_t = d(S_{1t}, S_{2t}) \qquad (2)$$

   where $d(\cdot)$ is the user-chosen distance function as in (1);

   c) repeat a) and b) T number of times, with T being a number typically ≥ 100;

3. Compute the difference F_{it} between each observed D_i and D_t, for t=1 to T, or

   $$F_{it} = D_i - D_t \qquad (3)$$

4. Record in a new vector V_i of length N the number of times out of T times that $D_i \leq D_t$, producing a value in [0, 1].

5. Additionally, for each V_i in Step 4, if it is greater than 0.95, a value of 1 is recorded in a new vector of length N, otherwise, 0 is recorded.

The result from Step 4 can be bimodal, depending on the random generation assumption, and the construction of a dichotomous variable in Step 5 can thus be useful. Either variables from Step 4 or 5 can be applied in further regression analysis. The variable from Step 5 can also be regarded as a test statistic because it is based on the 95% confidence that the observed distance for a given dyadic or triadic sequence set is no greater than randomly generated hypothetical distances. The computation of Steps 1 to 3 involves two loops, a loop of *T* times and a second loop of *N* times. An R package `linkedSeqs`, available upon request, is prepared for performing these computations. See Appendices A and B for the documentation for the two functions for computing degree of linked lives of dyadic and triadic sequence data.

The R packages contains two functions, `linked2seqs` and `linked3seqs`, for computing the degree of linked life course sequences and a few other associated statistics. In addition to supplying dyadic or triadic sequence data defined by `seqdef` of `TraMineR`, the user can choose a distance measure and its associated parameters, a random generation mechanism (see below), a random seed number (for starting the simulated dyadic or triadic data generation), and the number of simulated dyadic or triadic sequences. For `linked3seqs`, the user can additionally choose a weight for assigning to the distance between member 1 and 2, 1 and 3, and 2 and 3 of the triad, respectively.

Random Sequence Generation Mechanisms

I have included three random sequence generation mechanisms for performing Step 2.a as discussed below.

1. *Complete random generation*: By making this assumption, sequences of length *s* is randomly drawn from the observed alphabet (i.e., qualitative states) with replacement,

with no regard to the observed proportion of each state or to the meaningful order of states.

2. *State-conditional random generation*: By making this assumption, sequences of length *s* is randomly drawn from the observed alphabet (i.e., qualitative states) with replacement, with a probability proportional to the observed proportion of each state of the observed set of dyadic or triadic members (e.g., set of fathers' states, set of mothers' states, or set of children's states) but with no regard to the meaningful order of states.
3. *Sequence-conditional random generation*: By making this assumption, sequences of length *s* is randomly drawn from the observed set of dyadic or triadic members (e.g., set of fathers' sequences, set of mothers' sequences, or set of children's sequences). Using this assumption preserves the meaningful order of states and is useful when certain states cannot precede certain other states, such as divorce cannot precedes first marriage.

The choice of a random sequence generation mechanism depends on the nature of sequences and the substantive need of the research.

A Simulation Study

To assess the statistical properties of the proposed method introduced in the previous section, I conducted a simulation study of randomly generated sequences of 100 spells that belong to a dyadic data set of size *N*, containing Subset 1 and Subset 2 for the paired dyadic sequence data. The paired sequences are generated one pair at a time, and are entered into Subset 1 and Subset 2. Two sequences with no common state are maximally dissimilar (Dijkstra and Taris 1995; Elzinga 2003). For simulating the two subsets, I randomly generated, with replacement, a varying proportion of distinctive states. To test how the two members of a dyadic pair differ, I randomly assigned 100% of the alphabet of A, B, and C into a sequence of Subset 1, and randomly allocated 10%, 20%, …, 90% of the alphabet of D, E, and F into the counterpart sequence in Subset 2. That is, this allocated percentage increased from 10% to 90% by 10% increment. Note that 0% and 100% are trivial cases that need no simulation. At one extreme (10%), the paired sequences resemble each other little; at the other extreme (90%), the paired sequences resemble each other extremely well in terms of the alphabet contained. I performed this operation with *N*=100, 200, 500, 1,000, and 2,000 for the nine shared percentages. The simulation, using the procedure described in the previous section with *T*=1,000 for simulating *T* number of simulated dyads, is to be repeated *Z* times. (Note that for this version of the paper, I could complete only 30 paths of the simulation, due to its computational intensity. Ideally, *Z*=1,000 is desired, and will be attempted at a later time. However, there is no difference in the patterns between the results from just 1 path and 30 paths of the simulation.)

During each execution of the simulation situations (defined by sample size and percent shared states), the program `linked2seqs` (documented in Appendix A) is called. To focus on the difference of distinctive states (without the possible interference by differences in timing, duration, or order of states), I chose for running the program a particular distance measure, OMspell with an expansion cost of 0.5 and an indel cost of 2. This distance measure with the specified parameters is the most neutral in terms of sensitivity to timing, duration, and order (located nearest to the center of Figure 1 in Studer and Ritschard 2016). Using this distance measure in the simulation to minimize the influences of timing, duration, and order, the difference between two sets of sequences would be driven almost entirely by their distinctive states. Figure 1 presents the simulation results. The first random generation assumption is used because the simulated sequence data contain simply differing proportions of qualitative states, with no particular logical order of the states imposed.

---**Figure 1:** Boxplots Assessing Difference between “Observed” and 1,000 Simulated Dyadic Sequences—

The figure contains five panels of boxplots, each of which for a specific sample size, from 50 to 1,000 dyadic paired sequences. In each panel, nine sets of simulations are reported, from 10% to 90% shared states in the paired dyadic sequences. The plotted values are the output from the program `linked2seqs`, or the proportion of hypothetical or simulated dyadic sequences each of the current “observed” dyadic sequences outperforms, or having no greater distance than, the simulated dyadic sequences (the word “observed” is in quotes because in the simulation study, these “observed” dyadic sequences are also simulated).

We can make several general observations about the simulation results in Figure 1. First, the performance of the proposed method is not linear. That is, the proportion of simulated sequences outperformed does not correspond exactly to the proportion of shared states in the paired dyadic sequences. I will hold off offering a definitive explanation until a much large number of paths of the simulation is conducted. Second, the method is more sensitive to the changes in the middle range of percent shared states, from 50% to 70%, as seen in the current simulation results. The 60% of shared states provides the widest spread in the comparison of the “observed” and simulated dyadic sequences. Third and finally, the method is robust with regard to sample size. As is obvious from the figure, the shape of the distribution of the boxplots is almost identical across all sample sizes, with the larger sample sizes having a noticeable wider spread for those of 50% to 70% percent shared states.

I also plan to do a simulation of triadic sequences at a later time.

Empirical Application 1

In the first empirical application, I focus on the dyadic life course data from the U.S., using data from the Longitudinal Study of Generations (LSOG) as in Fasang and Raab's (2014) study. The LSOG sequences record family trajectories of middle-class parents born around 1920 to 1930 whose family formation took place roughly between 1935 and 1960 and their children whose family formation took place between 1955 and 1990.

The LSOG combines prospective and retrospective features, thus providing complete family formation sequences of parents and their children between ages 15 and 40. Fasang and Raab's (2014) research represents a first attempt to fully exploit the unique intergenerational and longitudinal information on family formation in the LSOG. Like them, I here use data for two generations: (1) the parent generation, the so-called silent generation born in the 1920s and 1930s, and (2) their children, the Baby Boom generation born in the late 1940s and 1950s. After data cleaning and preparation, they had 226 families with 342 parents and 305 children, resulting in 461 parent-child dyads in their study. Sixty-four families (or 98 parents) have more than one child. These parents appear in more than one parent-child dyad. On the other hand, there are 156 children forming both the mother-child dyad and the father-child dyad. Therefore, these children enter the analysis sample twice. The remaining children belong to the one parent-child dyad. The majority (66.4 %) of the single-parent dyads are mother-child dyads. For further details on the data, see Fasang and Raab (2014).

The dyadic sequence data contain nine family formation states: single, no child; single, one or more children; married, no child; married, one child; married, two children; married, three children; married, four or more children; divorced, no children; and divorced, one or more children. Note that there is a logical order to these nine family formations states. For example, the single state always happens first, and either of the two divorced states cannot precede any of the married states. Furthermore, the states with children are also ordered such that a state with a higher number of children always succeeds a state with a lower number of children or no child. Because of such logical orders of the sequence data, I chose random generation mechanism 3 for analyzing the data.

Another feature of such typical family formation sequences is the three distinctive characteristics of timing, duration, and sequence (order). They represent the onset of a particular state, the time a person spends in that particular state, and the sequencing of this and another state or other states, respectively. Because of these characteristics, I employed four different distance measures, with three of them more sensitive to one of the characteristics and the last one, equally insensitive to all three. These are the Hamming distance (for timing), the SVRspell distance (for duration), the OMspell distance with an expansion cost of 0.4 and a substitution matrix of INDELS (for order), and the OMspell distance with an expansion cost of 0.5 and an indel cost of 2 as the choice relatively neutral to timing, duration, and order. I analyzed the LSOG data of 461 dyadic sequence pairs, and present in Figure 2 the density plots of the degree of linked life courses according to these four dissimilarity measures.

---Figure 2: Density Plots of Degree of the LSOG Linked Life Courses, Using 1,000 Simulated Dyads by Random Generation Mechanism 3 (*N*=461)**—**

The X-axis measures the degree of linked life courses, or how much paired dyadic sequences are associated compared against randomly chosen sequences from the same subsets of data (i.e., the parent sequence or the child sequence subset). It appears that these LSOG dyads resemble each other more in terms of sequencing, followed by timing, then duration. The neutral distance measure (with a degree of linked life courses μ=0.547) appears to be a summary measure of the other three (timing focused, μ=0.569; duration focused, μ=0.515; and order focused, μ=0.591), locating itself among the other curves in a middle position. To visualize the actual sequence data, I present the LSOG sequence index plots in Figure 3, sorted by the neutral focused degree of linked life courses measure from above.

---Figure 3: Sequence Index Plots of the LSOG Parent-Child Dyads, Sorted by Degree of the Neutral Focused Linked Life Courses (*N*=461)—

The colored lines in the two plots at the same level are paired dyads. Therefore, it is obvious that those with a high degree of linked life courses gravitate toward the bottom of the plots, and those with a low degree, the top. Moving from the bottom to the top of the plots, we see that similarity between the dyadic members gradually decreases, until they are almost entirely dissimilar (other than the initial short duration of being single).

To apply the degree of linked life courses variable in a serious substantive study is beyond the scope of the current paper. However, to illustrate how this variable may be associated with some typical socioeconomic covariates, I stratify the analysis by children's education and present below the same set of density and sequence index plots as those above. Figure 4 contains two panels of density plots, the top panel for those children with at least a college education and the bottom, those without such an education.

---Figure 4: Density Plots of Degree of the LSOG Linked Life Courses by Children's Education, Using 1,000 Simulated Dyads by Random Generation Mechanism 3 (N_1=251; N_2=210)—

Two general comments are in order. First, those with higher education (N_2=210) tend to have a higher degree of linked sequencing or order. Their more linked life courses overall between the parent and the offspring generation is summarized by the purple curve representing the neutral focused approach (μ=0.598). On the other hand, the less educated (N_1=251), in comparison, tend to have less linked life courses overall, especially in terms of duration (the green curve), and even their neutral focused measure produced a much lower degree of linked life courses (μ=0.504). To visualize their actual life course sequences, I present the corresponding sequence index plots in Figure 5.

---Figure 5: Sequence Index Plots of the LSOG Parent-Child Dyads by Children's Education, Sorted by Degree of Linked Life Courses of the Neutral Focus (N_1=251; N_2=210)—

In addition to the greater degree of overall resemblance of the life course sequences among the dyads with highly educated children, another clear distinction is in the duration of certain states. For example, the dyads with less educated children in the bottom panels have much longer single duration among the children than their parents and much shorter married with children states. By contrast, the highly educated dyads have quite similar lengths in the single state between the two generations. These intergenerational educational effects are largely consistent with those reported in Fasang and Raab (2014) although their study also examined the question of educational mobility.

Empirical Application 2

To illustrate an analysis of the degree of linked life courses among triads, I use the data from the 2001-2015 waves of the U.S. Panel Study of Income Dynamics (PSID) and its Transition into Adulthood Supplement (TAS). The PSID, a longitudinal panel study, began in 1968, though complete labor force calendar data only exists for 2000 and afterward. The PSID collects information on all individuals in the household, but the most detailed employment information is only collected on the primary adult(s) in the household. This design might potentially omit detailed employment information for young adult household members who do not yet head their own households. To remedy this, the PSID also conducts the TAS, which collects more detailed information on topics including education and employment of respondents who were born between 1985 and 1997 *and* participated in the Child Development Supplement of the PSID. The TAS, first conducted in 2005 and designed as a supplement to the rich economic information collected on economically independent household heads, has been carried out biennially for those respondents who were aged over 18, were not attending high school at the time of the survey, had previously participated in a Child Development Supplement interview, and were part of a family that participated in the given year's main PSID interview. It contains both young adults living as heads of households and those who still live with their parents or other PSID heads of households.

The PSID contains information on a wide variety of topics, but its employment history data are analysed in this paper. The PSID also allows for linkage of monthly employment histories between children and their parents because of the household panel design. Furthermore, the PSID asks whether respondents were out of the labor force or unemployed for at least one week during each calendar month, but the exact start or end dates of employment spells are not given. Respondents who report being unemployed for at least a week during the month are considered unemployed for the entire unit. Likewise, respondents who reported being out of the labor force are similarly considered though priority is given to unemployment if respondents report both states in the same spell.

The PSID sample analyzed covers a period from January 2000 to December 2014, corresponding to the first 30 months of the parents' data when the children were early teenagers and the last 30 months of the children's data when they were eligible for labor force participation. I applied the proposed method to the PSID data for computing the degree of linked life courses among the triadic sequence data.

Because there is no particular logical order to the sequencing of three "employment," "unemployed," and "out of labor force" states, I applied random generation mechanisms 1 and 2. The distance measure of choice is OMloc with an expansion cost of 0.4 (Figure 1, Studer and Ritschard 2016) as the main concern here is with the duration of states. In addition, weighting the resulting distances between sequences can be necessary because we now have triadic sequence data. Each triad in the PISD data contains two parents and a child. The distance between the parents, between father and child, and between mother and child can be weighted differentially in the calculation by using W1, W2, and W3. I first applied equal weights of W1=W2=W3 before using a set of unequal weights of 2W1=W2=W3, where the weight of W1, the weight for the distance between the two parents, is half as important as W2 or W3, the weight for the distance between a parent and the child. Figure 6 reports the density plots of the degree of linked life courses in the PSID triads.

---Figure 6: Density Plots of Degree of Linked Life Courses, PSID, Using 1,000 Simulated Triads by Random Generation Mechanisms 1 and 2 with Equal Weights and Unequal Weights (*N*=439)—

Random generation mechanism 1 of complete randomness tends to produce a lower degree of linked lives measure, equally weighted or unequally weighted. In contrast, the conditional random assumption of generation mechanism 2 tends to result in a higher degree of linked life courses, especially by assigning a weight to the parent-child distance twice of that between the two parents (the blue curve). Because the conditional random assumption as well as the unequal weighting is more reasonable, I generated sequence index plots for the triadic sequence data (Figure 7).

---**Figure 7:** Sequence Index Plots of Parent-Child Dyads, Sorted by Degree of Linked Life Courses of the Duration Focus (*N*=439)—

Once again, each thin line going through the three panels represents a particular triad. Toward the lower end of the plots we find those triads with all three members employed the entire 30 months, though in different years for the parents versus the children. Moving to the top, we find gradually more triads whose members' life courses are not linked, showing different durations, different states, or both. Thus, parents' and children's employed, unemployed and out of labor force spells tend to be rather linked. It is obvious from the plots as well that overall parents and children's lives are interdependent or linked to a good degree. The measure of linked life courses shown here in the figure has a mean of 0.750.

As in the first empirical application, I continue the analysis one step further by adding race as a covariate. This is a dichotomous variable containing the two categories of "whites" and "nonwhites." Figure 8 presents density plots of the degree of linked life courses measure of the four scenarios by assuming random generation mechanisms 1 and 2 with both a set of equal as well as a set of unequal weights, as previously defined.

---**Figure 8:** Density Plots of Degree of Linked Life Courses by Race, PSID, Using 1,000 Simulated Triads by Random Generation Mechanisms 1 and 2 with Equal Weights and Unequal Weights (N_1=271; N_2=168)—

As in Figure 6, the curves based on random generation mechanism 1 is flatter than those based on random generation mechanism 2. The two density curves based on random generation mechanism 2 (red and blue) indicate a much higher degree of linked life courses among whites than among nonwhites. As before, I use the results based on random generation mechanism 2 with unequal weights to produce further sequence index plots (Figure 9).

--- Figure 9: Sequence Index Plots of Parent-Child Dyads by Race, PSID, Sorted by Degree of Linked Life Courses of the Duration Focus (N_1=271; N_2=168): Top Row: Whites; Bottom Row: Nonwhites—

Judged by the sequence index plots, we can conclude that the white triads have a much higher degree of linked lives than their nonwhite counterparts. This is indicated primarily by the much greater area of shared employed spells, and, to a smaller degree, by the relative correspondence among the members of the other two states as well, with an overall mean of the degree of linked lives measure of 0.796. In contrast, the association of the state spells is noticeably weaker for the nonwhite triads, with the mean of the degree of linked lives measure of 0.675, an 18% difference from their white counterparts.

Future Directions

I do not plan to draw further general conclusions about the proposed method because its performance is as expected and because its interpretation is straightforward even though further simulation is yet to be conducted. Instead, I would like to offer some brief reflections on possible extensions to the current paper.

First, as mentioned in the introduction, the proposed method of linked life courses has the flexibility of extending to the analysis of tetrads, pentads, and hexads, or even of higher dimensional polyads. How can this goal be accomplished? The main procedure discussed in an earlier section should still apply. The only place that needs further adjustments is how distance weighting is handled. Just like with triads, unequal weighting can be appropriate, depending on the actual substantive meaning of the relationships between members of the polyad. However, the number of weights increases with the number of members in a polyad in the series of 1, 3, 6, 10, 15, 21, etc. Similarly, the difficulty with assigning meaningful unequal weights also increases.

Another extension of the proposed method is in its application. Obviously, the degree of linked life course measure can be used as an outcome variable or as an explanatory variable in substantive analyses. This would amount to extending the stratified analysis by education or by race to a full-blown analysis with a meaningful number of covariates. Another related possibility, as in the situations where random generation mechanism 1 or 2 is reasonable, is that we can further analyze the degree of linked lives as a dichotomous variable even without going through a cluster analysis, as suggested by Figures 6 and 8, as long as we make a decision about where the cut-off point is, a decision related to the next issue.

An interesting question to consider is to what degree a polyad's life courses can be regarded as linked. This is really up to the researcher to decide. If one is rigorous about the selection criterion, one can use the 95% rule, that is, an observed polyad should not have distances greater than the distances for those for the 95% of the simulated polyadic counterparts. The current R package outputs this result as a byproduct. The question that I have not considered is to what degree a polyad's life courses can be regarded as not linked at all. This, again, is an issue left for the data analyst.

References

Abbott, Andrew, and Forrest, John. 1986. "Optimal Matching Methods for Historical Sequences." *Journal of Interdisciplinary History*, **XVI**, 3, 471-494.

Aisenbrey, Silke, and Anette E. Fasang. 2017. “The Interplay of Work and Family Trajectories over the Life Course: Germany and the United States in Comparison.” *American Journal of Sociology 122*(5): 1448-1484.

Amory, C., et al. 2015. “Comparison between Observed and Simulated Aeolian Snow Mass Fluxes in Adélie Land, East Antarctica.” *The Cryosphere* 9: 1373-1383.

Barban, Nicola, et al. “Causal Effects of the Timing of Life-course Events: Age at Retirement and Subsequent Health.” *Sociological Methods & Research* (2017): 0049124117729697.

Bengtson, Vern L., Glen H. Elder, Jr., and Norella M. Putney. 2005. “The lifecourse Perspective on Aging: Linked Lives, Timing, and History.” Pp. 493-509 in *The Cambridge Handbook of Age and Ageing*, edited by Malcolm L. Johnson. New York, NY: Cambridge University Press.

Blanchard, Philippe, Felix Bühlmann and Jacques-Antoine Gauthier (eds). 2014. *Advances in Sequence Analysis: Methods, Theories and Applications*. New York: Spinger.

Brinberg, Miriam, et al. 2016. “Analyzing Dyadic Data Using Grid-Sequence Analysis: Interdyad Differences in Intradyad Dynamics.” *Journal of Gerontology Series B* 73: 5-18.

Conger, Rand D. and Glen H. Elder, Jr. 1994. *Families in Troubled Times: Adapting to Change in Rural America*. New York, NY: Aldine de Gruyter.

Cornwell, Benjamin. 2015. *Social sequence analysis: Methods and applications* (Vol. 37). Cambridge University Press.

Dijkstra, W., & Taris, T. 1995. Measuring the agreement between sequences. *Sociological methods & research*, *24*(2), 214-231.

Elder, Glen H., Jr. 1995. “The life Course Paradigm: Social change and Individual Development. Pp. 101-139 in *Examining lives in context: Perspectives on the Ecology of Human*

Development, edited by Phyllis Moen, Glen H. Elder, Jr., and Kurt Luscher. Washington, DC: American Psychological Association.

Elder, Glen J. Jr, Monica Kirkpatrick Johnson, and Robert Crosnoe. 2003. “The Emergence and Development of Life Course Theory.” Pp. 3-19 in *Handbook of the Life Course*, edited by Jaylen T. Mortimer and Michael J. Shanahan, New York, Plenum.

Elzinga, Cees H. 2003. Sequence similarity: a nonaligning technique. *Sociological methods & research*, *32*(1), 3-29.

Fasang, Anette E. and Tim F. Liao. 2014. “Visualizing Sequences in the Social Sciences: Relative Frequency Sequence Plots.” *Sociological Methods & Research* 43(4): 643-676.

Fasang, Anette Eva, and Marcel Raab. 2014. “Beyond Transition: Intergenerational Patterns of Family Formation among Middle-Class American Families.” *Demography* 51: 1703-1728.

Furman, Bradley T. et al. 2018. “Braun-Blanquet Data in ANOVA Designs: Comparisons with Percent Cover and Transformations Using Simulated Data.” *Marine Ecology Progress Series* 597: 13-22.

Gauthier, Jacques-Antoine, Eric D. Widmer, Philipp Bucher, and Cedric Notredame. 2010. “Multichannel Sequence Analysis Applied to Social Science Data.” *Sociological Methodology* 40: 1-38.

Gilligan, Megan, Amelia Karraker, and Angelica Jasper. 2018. “Linked Lives and Cumulative Inequality: A Multigenerational Family Life Course Framework.” *Journal of Family Theory & Review* 10: 111-125.

Liao, Tim Futing. 2002/2011. *Statistical Group Comparison*. New York: Wiley/Wiley Online Library (https://onlinelibrary.wiley.com/doi/book/10.1002/9781118204214).

Liefbroer, Aart C., and Cees H. Elzinga. 2012. “Intergenerational Transmission of behavioural Patterns: How Similar Are Parents’ and Children’s Demographic Trajectories?” *Advances in Life Course Research* 17: 1-10.

Macmillan, Ross and Ronda Copher. 2005. “Families in the Life Course: Interdependency of Roles, Role Configurations, and Pathways.” *Journal of Marriage and Family 67*(4): 858–879 (https://doi.org/10.1111/j.1741-3737.2005.00180.x)

Nightingale, Glenna. 2016. “R Package lifecourse.” The CRAN R project repository.

Raab, Marcel, Anette Eva Fasang, Aleksi Karhula, and Jani Erola. 2014. “Sibling similarity in family formation.” *Demography* 51(6): 2127-2154.

Studer, Matthias. 2013. WeightedCluster library manual: A practical guide to creating typologies of trajectories in the social sciences with R.

Studer, Matthias, and Ritschard, Gilbert. 2016. “What matters in differences between life trajectories: a comparative review of sequence dissimilarity measures.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 179(2): 481-511.

Studer, Matthias, Struffolino, Emanuela, and Fasang, Anette E. 2018. “Estimating the Relationship between Time-varying Covariates and Trajectories: The Sequence Analysis Multistate Model Procedure.” *Sociological Methodology* 0081175017747122.

Appendix A

linked2seqs {linkedSeqs} R Documentation

Degree of linked dyadic sequences

Description

The function computes the degree of linkedness of dyadic sequence data against random sequences, assuming one of the three random sequence generation mechanisms. It returns a list of five objects for user to perform further analysis of linked lives.

Usage

```
linked2seqs(p1.seq, p2.seq, a=1, m="HAM", sm=NULL, expcost=0.5,
indel=1.0, tpow=1.0, s=36963, T=1000)
```

Arguments

`p1.seq`	First State Sequence Object generated by the `seqdef` function of the TraMineR package.
`p2.seq`	Second State Sequence Object generated by the `seqdef` function of the TraMineR package.
`a`	Integer. Default 1. Random sequence generation mechanism. 1: Random pairs of sequences generated using just the alphabet. 2: Random pairs of sequences generated using the alphabet proportional to *p1.seq* and to *p2.seq*, respectively. 3: Random pairs of sequences generated using the observed *p1.seq* and *p2.seq* sequence data.
`m`	String. Default "HAM". Method for computing sequence distances. See documentation for `seqdist`.
`sm`	Matrix, Array, or String. Default NULL. Substitution matrix. See documentation for `seqdist`.
`expcost`	Double. Default 0.5. The cost of spell length transformation. See documentation for `seqdist`.
`indel`	Double. Default 1.0. Insertion/deletion cost. See documentation for `seqdist`.
`tpow`	Double. Default 1.0. The exponential weight of spell length. See documentation for `seqdist`.
`s`	Integer. Default 36963. Seed for random sequence generation.
`T`	Integer. Default 1000. Number of random sequences to compute.

Value

The function outputs five objects, *prop.LinkedSeqs*, *degree.linkedSeqs*, *sig.linked.Seqs*, *observed.dist*, and *random.dist*.

`prop.LinkedSeqs`	This scalar in [0,1] gives the overall proportion of linkedness in the entire dataset, based on *degree.linkedSeqs*.
`degree.linkedSeqs`	This [0,1] variable defines for each pair of sequences the proportion of randomly generated sequences the observed pair outperforms, i.e., with a distance no greater than the randomly generated distances.
`sig.linked.Seqs`	This 0-1 variable recodes *prop.LinkedSeqs* by assigning a value of "1" if the observed distance outperforms 95 percent of the random distances for the dyad, "0" otherwise.

`observed.dist`	It contains the observed distances between the paired data.
`random.dist`	It contains the random distances for all the random pairs.

Examples

```
dyads.results1 <- linked2seqs(ma.seq,child.seq,a=2,s=123,T=10000)
dyads.results2 <-
linked2seqs(sib1.seq,sib2.seq,a=3,m="OM",sm="CONSTANT",s=123,T=50000)
```

[Package *linkedSeqs* version 1.0.0 Index]

Appendix B

linked3seqs {linkedSeqs} R Documentation

Degree of linked sequences

Description

The function computes the degree of linkedness of triadic sequence data against random sequences, assuming one of the three random sequence generation mechanisms. It returns a list of five objects for user to perform further analysis of linked lives.

Usage

```
linked3seqs(p1.seq, p2.seq, p3.seq, a=1, m="HAM", sm=NULL, expcost=0.5,
indel=1.0, tpow=1.0, w=c(1,1,1), s=36963, T=1000)
```

Arguments

`p1.seq`	First State Sequence Object generated by the `seqdef` function of the TraMineR package.
`p2.seq`	Second State Sequence Object generated by the `seqdef` function of the TraMineR package.
`p3.seq`	Third State Sequence Object generated by the `seqdef` function of the TraMineR package.
`a`	Integer. Default 1. Random sequence generation mechanism. 1: Random triads of sequences generated using just the alphabet. 2: Random triads of sequences generated using the alphabet proportional to *p1.seq* and *p2.seq* together and to *p3.seq*, respectively. 3: Random triads of sequences generated using the observed *p1.seq*, *p2.seq*, and *p3.seq* sequence data.
`m`	String. Default "HAM". Method for computing sequence distances. see documentation for `seqdist`.
`sm`	String. Default NULL. Substitution matrix. see documentation for `seqdist`.
`expcost`	Double. Default 0.5. The cost of spell length transformation. See documentation for `seqdist`.
`indel`	Double. Default 1.0. Insertion/deletion cost. See documentation for `seqdist`.
`tpow`	Double. Default 1.0. The exponential weight of spell length. See documentation for `seqdist`.
`w`	Vector. Default c(1,1,1). The weights applied to the distances between the first and the second, the first and the third, and the second and the third members of the triad.
`s`	Integer. Default 36963. Seed for random sequence generation.
`T`	Integer. Default 1000. Number of random sequences to compute.

Value

The function outputs five objects, *prop.LinkedSeqs*, *degree.linkedSeqs*, *sig.linked.Seqs*, *observed.dist*, and *random.dist*.

`prop.LinkedSeqs`	This scalar in [0,1] gives the overall proportion of linkedness in the entire dataset, based on *degree.linkedSeqs*.
`degree.linkedSeqs`	This [0,1] variable defines for each triad of sequences the proportion of randomly generated sequences the observed triad outperforms, i.e., with a distance no greater than

	the randomly generated distances.
`sig.linked.Seqs`	This 0-1 variable recodes *prop.LinkedSeqs* by assigning a value of "1" if the observed distance outperforms 95 percent of the random distances for the triad, "0" otherwise.
`observed.dist`	It contains the average observed distances between the triadic sequences.
`random.dist`	It contains the average random distances for all the random triads.

Examples

```
triads.results1 <- linked3seqs(pa.seq,ma.seq,child.seq,a=2,s=123,T=10000)
triads.results2 <- linked3seqs(sib1.seq,sib2.seq,sib3.seq,a=3,s=123,T=50000)
triads.results3 <-
linked3seqs(pa.seq,ma.seq,child.seq,a=3,m="OM",sm="CONSTANT",w=c(2,0.5,1),s=1
23,T=10000)
```

[Package *linkedSeqs* version 1.0.0 Index]

Figure 1: Boxplots Assessing Difference between "Observed" and 1,000 Simulated Dyadic Sequences with Spell Length=100 and Sample Size=50, 100, 200, 500, and 1,000 for Percent Shared States=10% to 90%, 30 Paths of the Simulation

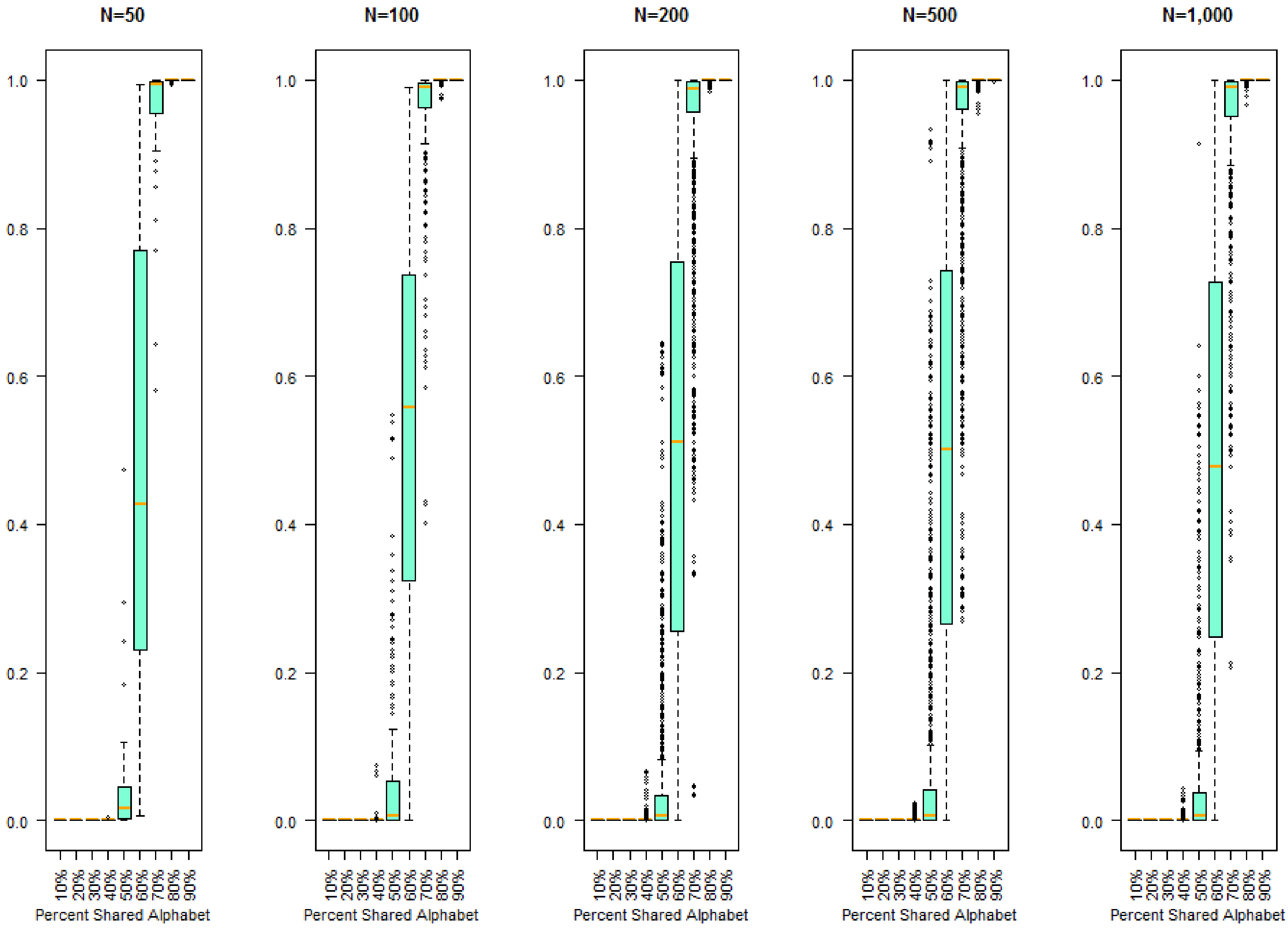

Figure 2: Density Plots of Degree of the LSOG Linked Life Courses, Using 1,000 simulated Dyads by Random Generation Mechanism 3 (*N*=461)

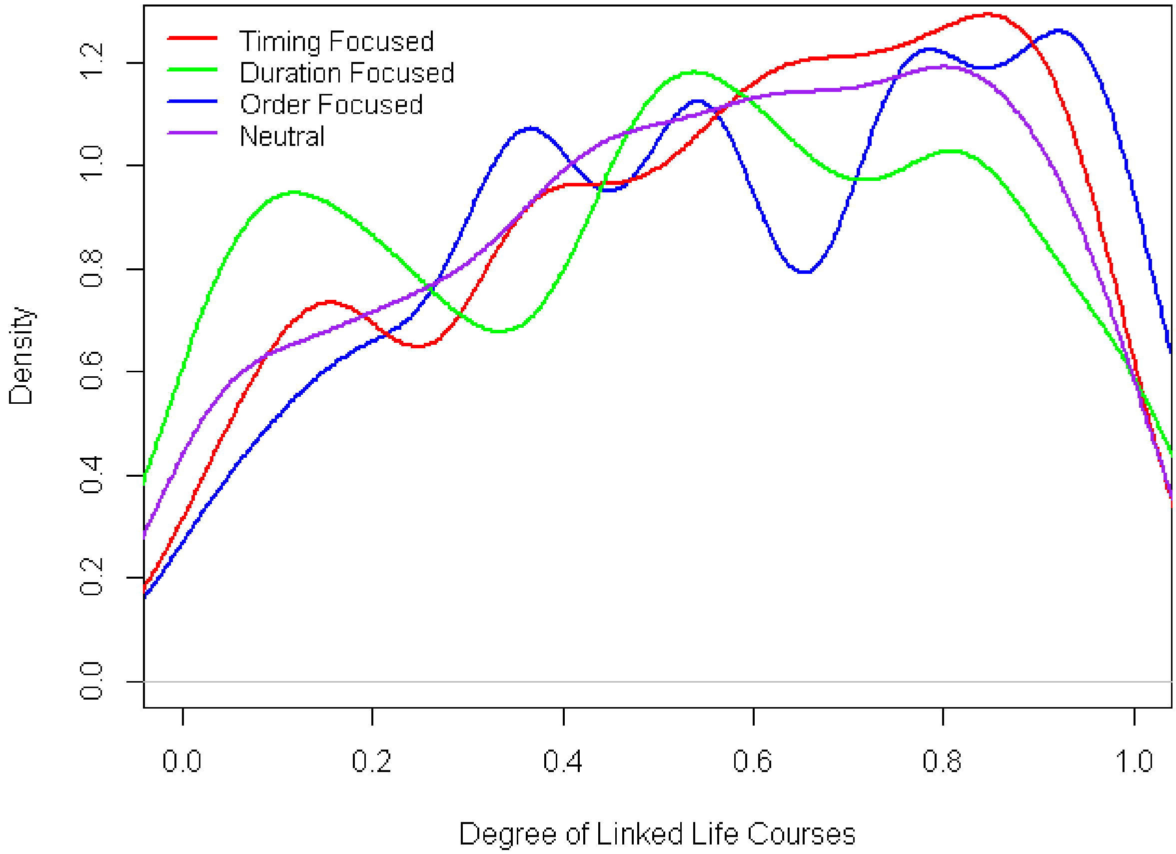

Figure 3: Sequence Index Plots of the LSOG Parent-Child Dyads, Sorted by Degree of the Neutral Focused Linked Life Courses (*N*=461)

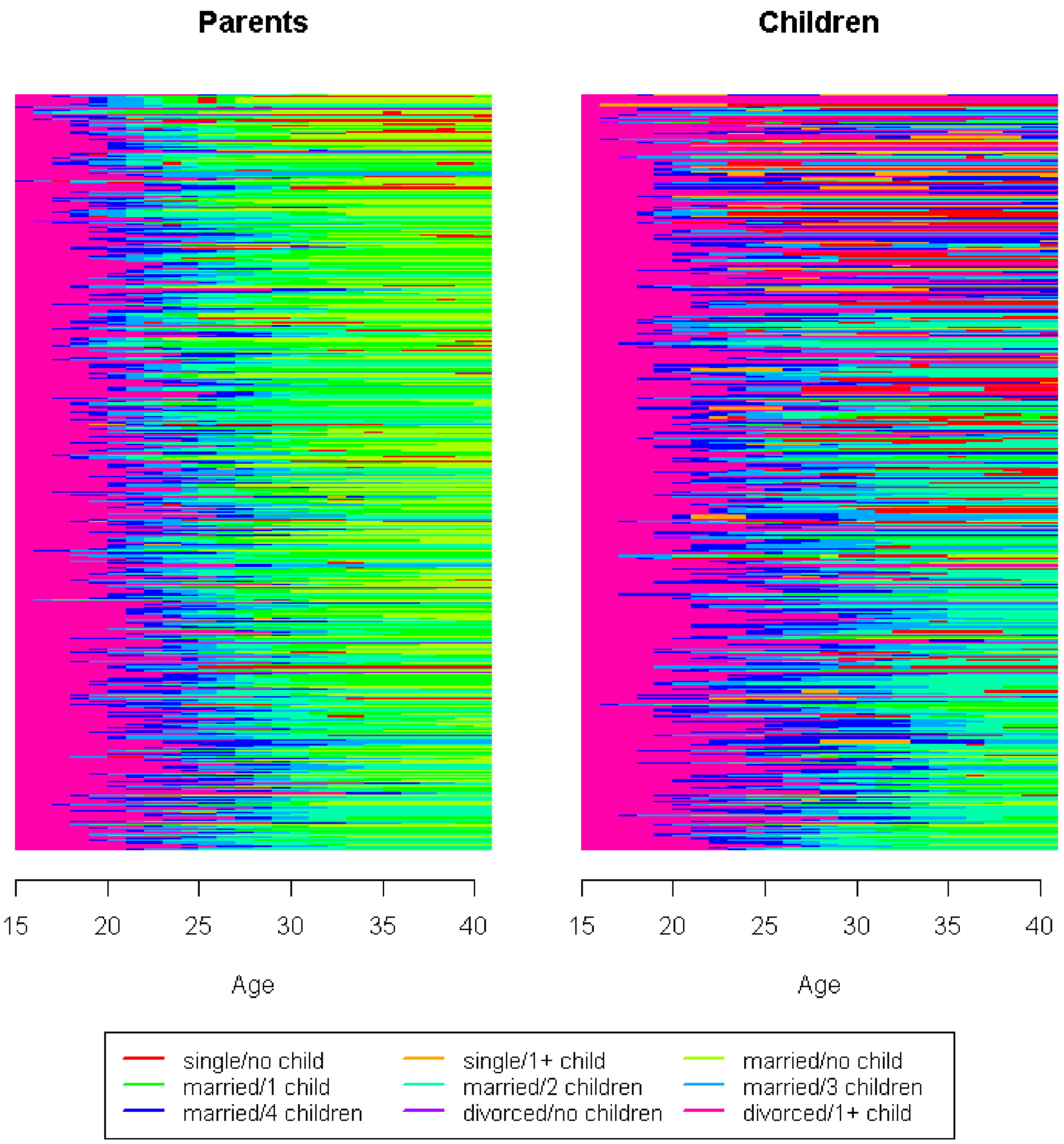

Figure 4: Density Plots of Degree of the LSOG Linked Life Courses by Children Education, Using 1,000 Simulated Dyads by Random Generation Mechanism 3 (N_1=251; N_2=210)

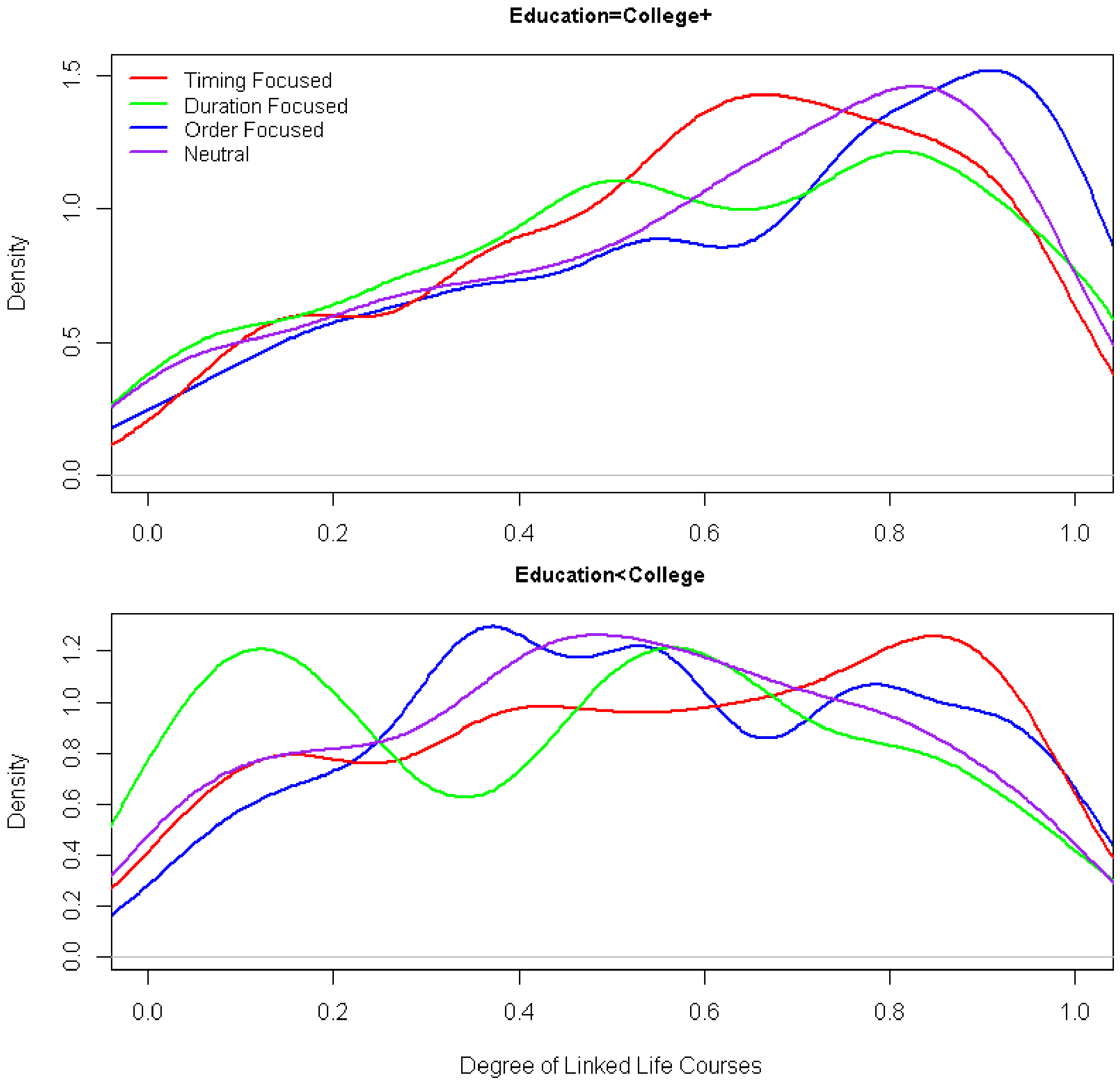

Figure 5: Sequence Index Plots of the LSOG Parent-Child Dyads by Children's Education, Sorted by Degree of Linked Life Courses of the Neutral Focus (N_1=251; N_2=210): Top Row: Education ≥ College; Bottom Row: Education ≤ College

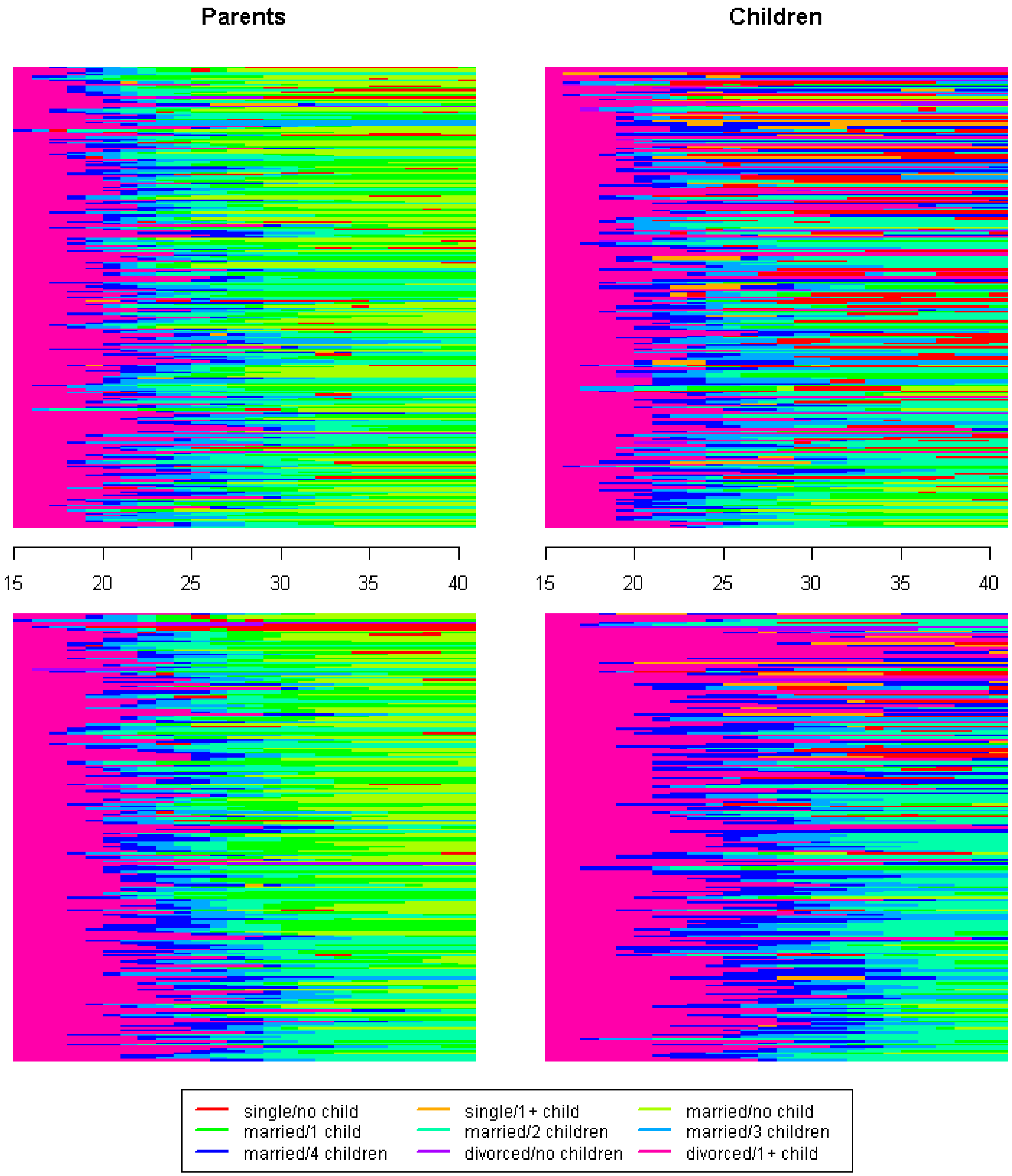

Figure 6: Density Plots of Degree of Linked Life Courses, PSID, Using 1,000 Simulated Triads by Random Generation Mechanisms 1 and 2 with Equal Weights and Unequal Weights (*N*=439)

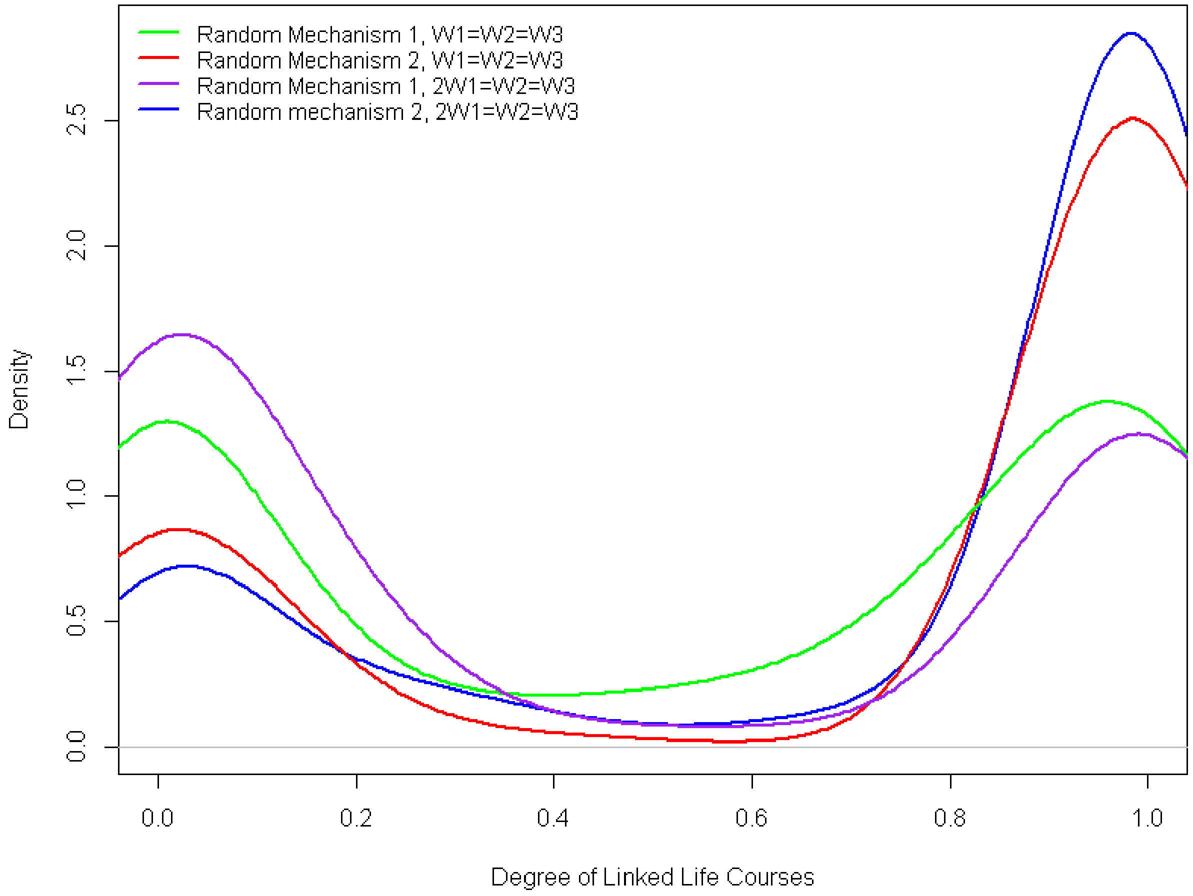

Figure 7: Sequence Index Plots of Parent-Child Dyads, Sorted by Degree of Linked Life Courses of the Duration Focus (*N*=439)

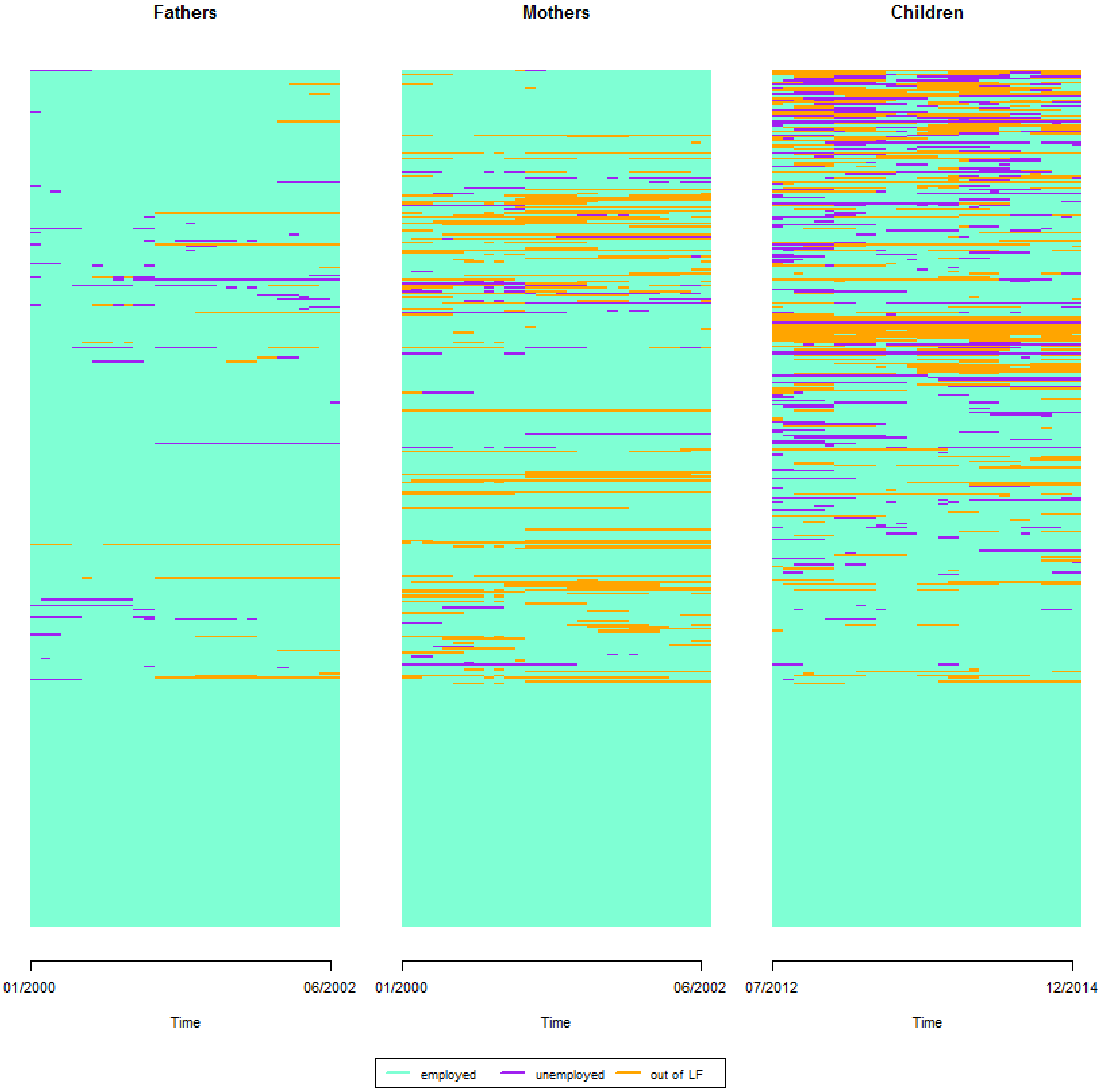

Figure 8: Density Plots of Degree of Linked Life Courses by Race, PSID, Using 1,000 Simulated Triads by Random Generation Mechanisms 1 and 2 with Equal Weights and Unequal Weights (N_1=271; N_2=168)

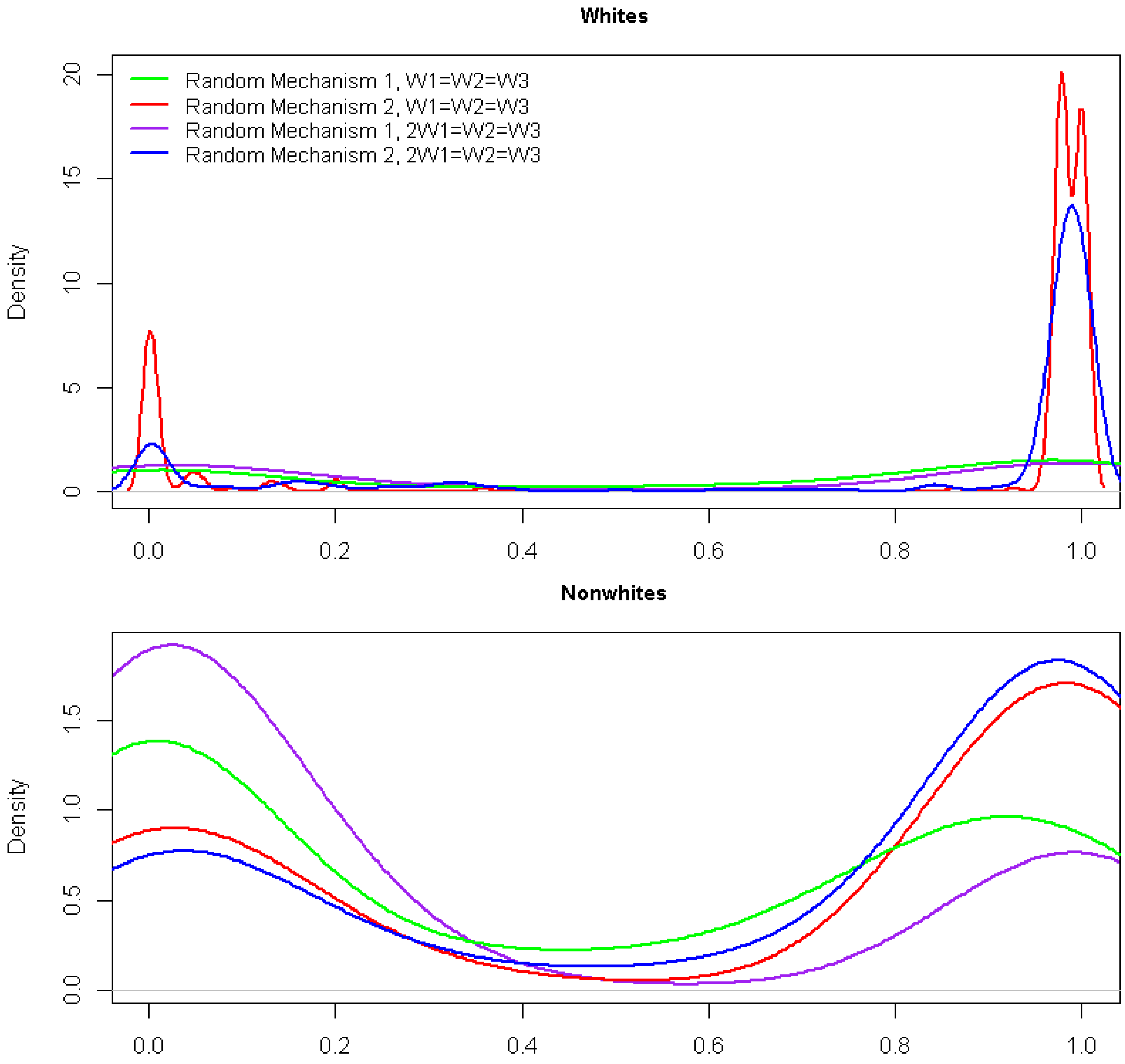

Figure 9: Sequence Index Plots of Parent-Child Dyads by Race, PSID, Sorted by Degree of Linked Life Courses of the Duration Focus (N_1=271; N_2=168): Top Row: Whites; Bottom Row: Nonwhites