Partnership, fertility, and employment trajectories of immigrants in the UK: An intersectional life course approach using three-channel sequence analysis

Supplementary Analyses

1. Robustness checks

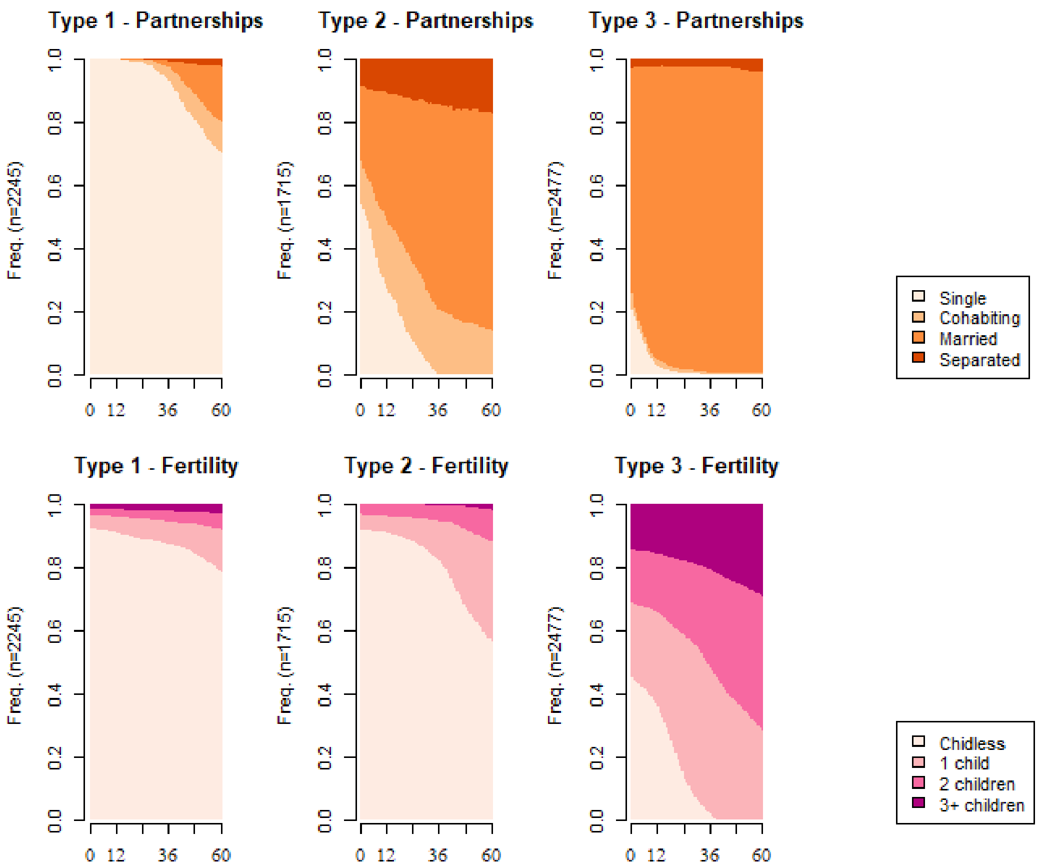
We have conducted a range of robustness checks to test the sensitivity of our results to: a) sample size limitations, b) the length of the observation window, c) the choice of cost regimes, distance measures, and clustering methods, and d) the number of selected clusters. We also discuss the validity of the MCSA approach and the sensitivity of our results to using weights during cluster analysis. We briefly discuss the results of these additional analyses below.

1.1 Sample size

Our analytical sample has reduced from 7,890 to 3,301 because employment histories were only collected for a subset of individuals in the UKHLS. This large reduction in sample size raises potential issues of selectivity. To check the robustness of our results to this sample size limitation, we have conducted two sets of additional analyses. First, we have conducted a two-channel sequence analysis using information only on immigrants' partnership and fertility histories (Figure S1). The sample size for this analysis is considerably larger, 6,437 individuals (we had to drop 1,453 individuals who did not have complete partnership and fertility histories for the five-year observation period).

There are some small differences between how individuals are classified into groups based on their partnership and fertility sequences in the three- and two-channel classifications. For example, in the two-channel analysis, a somewhat larger share of individuals with children are classified into the first and second group than in the three-channel analysis. Additionally, in the two-channel analysis, all individuals in the second and third clusters form a relationship 1-3 years after migration, whereas in the three-channel analysis some individuals remain never partnered in these groups. These small differences notwithstanding, in the two-channel analysis we find three partnership and fertility clusters, which correspond to those presented in the results section: 1) single and childless, 2) partnered and childless, and 3) family migration. These results indicate that the partnership and fertility groupings that emerge in the three-channel analysis are reliable, and our results are unlikely to be biased because of the limited size of the analytical sample.

Figure S1: Results of cluster analysis following a two-channel (partnership and fertility) sequence analysis (n = 6,437)



Source: Authors' calculations using UKHLS data, waves 1–9 (2009–2019).
Notes: The 'Separated' category includes separated as well as widowed individuals. The 'Single' category refers to never partnered individuals. Those who become single following union dissolution are included in the 'Separated' category. Those who repartner following separation or widowhood are included in the cohabiting or married category depending on the type of their partnership.

Table S1: Analytical sample and original sample (without sample selection for the analysis) by covariates

	Analytical sample		Original sample	
	N	%	N	%
Migrant origin				
EU	453	18.2	1,418	18.0
Rest of Europe & West	183	7.4	535	6.8
India	345	13.9	1,104	14.0
Pakistan	210	8.5	905	11.5
Bangladesh	159	6.4	562	7.1
Caribbean	115	4.6	277	3.5
Africa	597	24.0	1,797	22.8
Other	423	17.0	1,288	16.3
Migration cohort				
1958–1989	657	26.4	1,459	18.5
1990–2004	1,295	52.1	3,475	44.1
2005–2014	533	21.5	2,952	37.4
Age at arrival				
18–24	1,096	44.1	3,297	41.8
25–29	693	27.9	2,235	28.3
30+	696	28.0	2,354	29.9
Parental SES				
Professional & Managerial	843	33.9	2,332	29.6
Intermediate	825	33.2	2,576	32.7
Routine	576	23.2	1,810	23.0
Not working/deceased/absent	241	9.7	1,168	14.8
Sex				
Male	1,045	42.1	3,601	45.7
Female	1,440	57.9	4,285	54.3
Total	2,485	100	7,886	100

Source: Authors' calculations using UKHLS data, waves 1–9 (2009–2019).

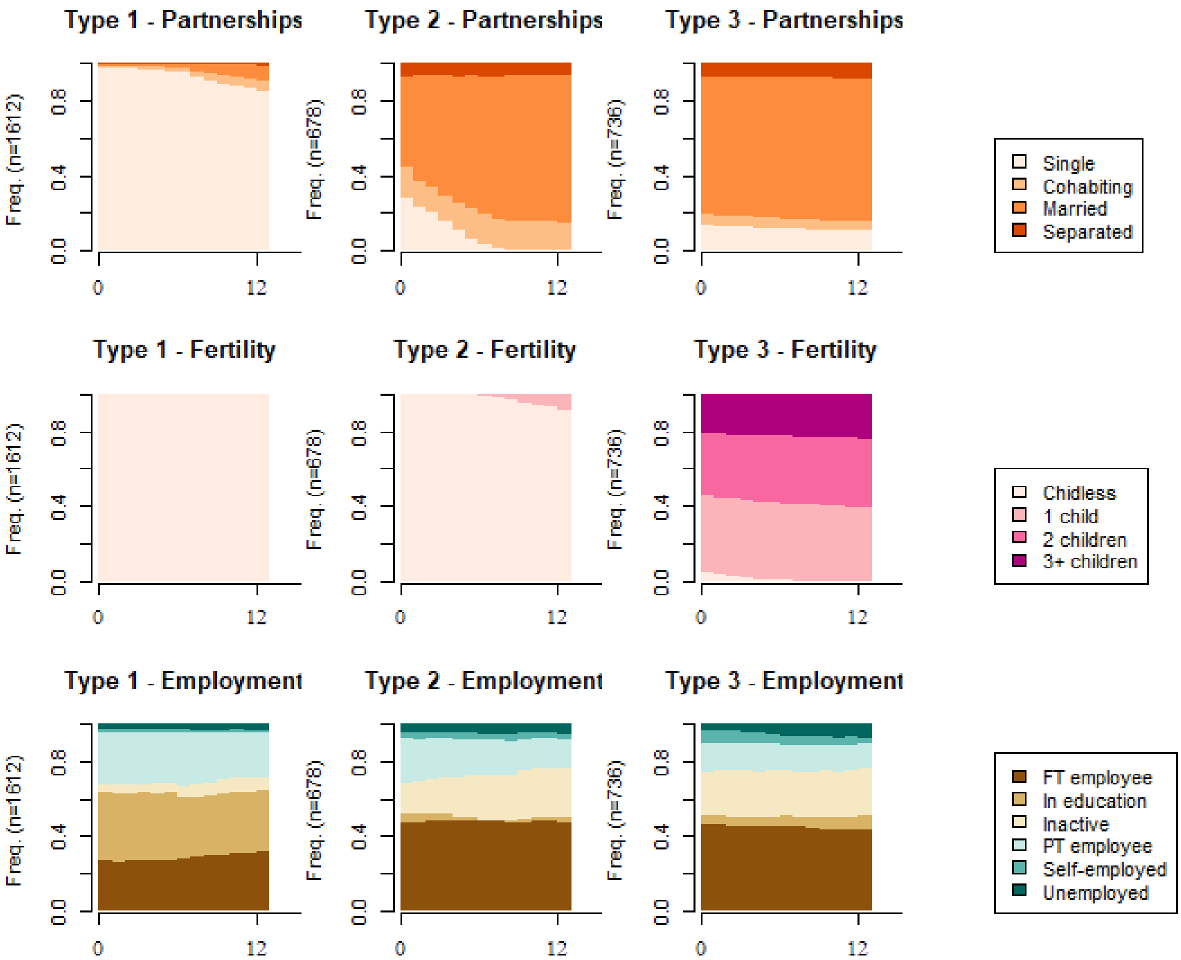
Second, we have compared the characteristics of the original sample and the analytical sample with respect to the variables used in the analyses (Table S1). The composition of the two samples is very similar regarding migrant origin, age at arrival, and sex. The analytical sample has a larger share of individuals from the two earlier migration cohorts and a smaller share of those who have arrived more recently (21% vs. 37% in the original data). This is partly because some more recent migrants were not yet in the UK when employment histories were collected (in waves 1 (Jan 2009 – March 2011) and 6 (Jan 2014 – May 2016)) and partly because due to their more recent arrival, fewer of the individuals in the 2005 – 2014 migration cohort could have been followed for 5 years. Similarly, although the share of those whose parents were in intermediate or routine occupations is almost identical in the two samples, the analytical sample has a somewhat larger share of those whose parents were in professional and managerial occupations and a somewhat lower share of those where the parents were not working. This is because a somewhat larger share of those whose parents were in professional and managerial occupations and a somewhat lower share of those whose parents were not working have employment histories. Taken together, we conclude that the analytical sample is very similar to the original sample and thus using the analytical sample is unlikely to lead to biased results.

1.2 Length of the observation window

We follow the evolution of immigrants' trajectories of partnership, fertility, and employment for five years after their arrival in the UK. Sequence analysis requires that all individuals have complete information for the duration of the analysis. This implies that individuals who are observed for less than five years after migration have to be excluded from the analysis. For our analysis this meant that a further 816 individuals had to be dropped. Choosing a shorter observation window would result in a larger analytical sample; the disadvantage is that not many individuals may experience partnership, fertility, or employment transitions soon after arriving in the UK.

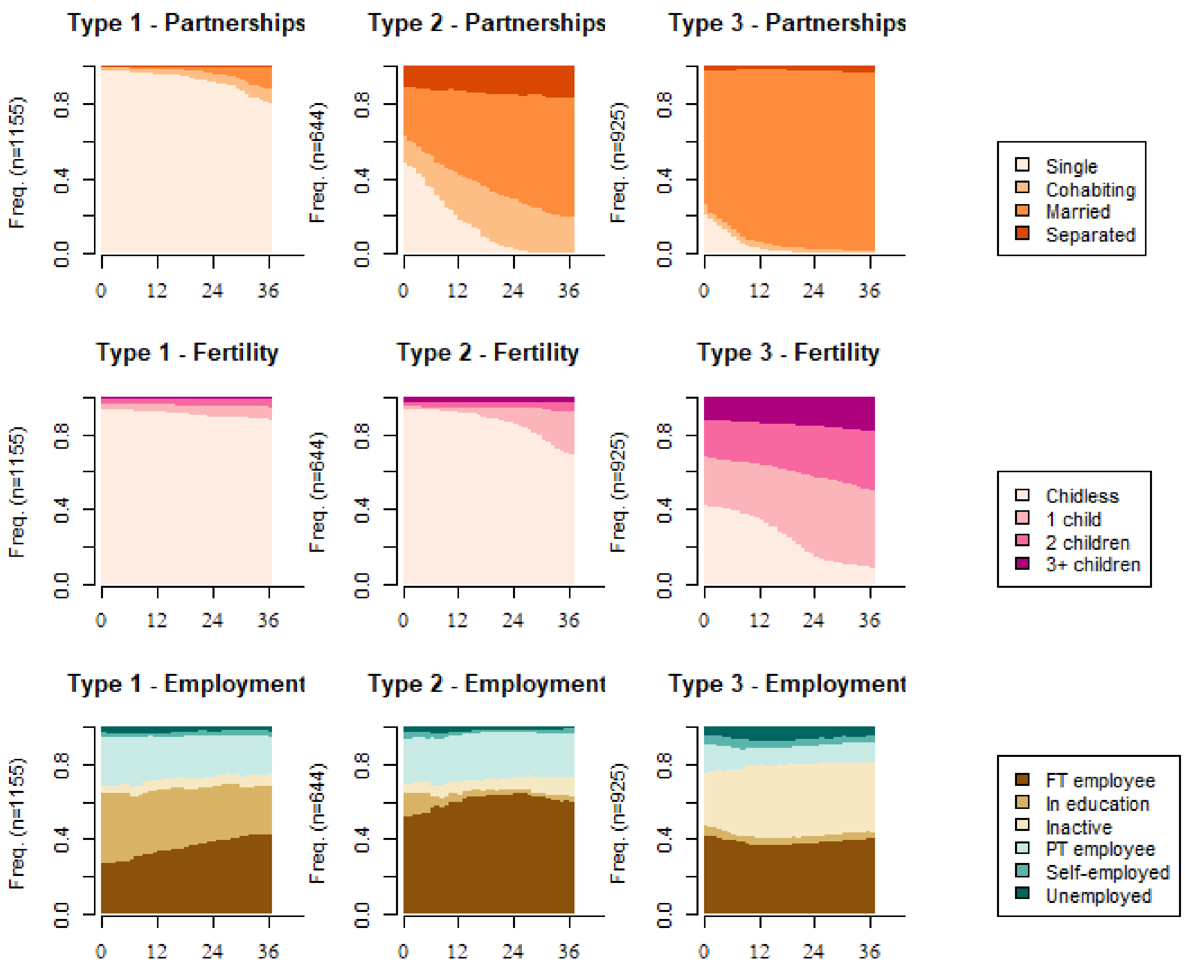
To test the sensitivity of our results to the length of the observation window, we have replicated the MCSA and cluster analysis using a one- (Figure S2), three- (Figure S3), and ten-year (Figure S4) observation window following Mikolai and Kulu (2019). Broadly speaking, we observe similar clusters regardless of the length of the observation window. However, only observing immigrants for one or three years after migration means that we do not yet observe many partnership, fertility, and employment transitions and thus, the clusters are mainly formed using information on immigrants' partnership and employment status as well as parity at the time of their arrival in the UK. This is reflected in the size of the first cluster (single, childless, students), which is the largest in both cases. As expected, we observe somewhat more transitions when we follow individuals for three years rather than only for one year. When following immigrants for 10 years after arrival in the UK, we see very similar results to what is presented in the paper. When observing individuals for a longer period, a larger share forms partnerships (primarily marriages) and have children, and more students become full-time employed over time, as would be expected. In the family migrants cluster, an even larger share becomes inactive over time. We draw two main conclusions from this set of robustness checks. First, although observing individuals for a longer period results in a smaller analytical sample, many transitions only occur later justifying our choice to follow individuals for five years after migration. We do not follow individuals for longer than five years as this would result in an even smaller sample and would lead to the same substantive findings and conclusions as what is presented in the main analysis. Second, regardless of the length of the observation window, we find similar groups in the data. This indicates that individuals who are dropped due to incomplete information for the entire five-year observation window are likely to have similar experiences to those who are retained for analysis.

Figure S2: Results of cluster analysis using a one-year observation window (n = 3,026)



Source: Authors' calculations using UKHLS data, waves 1–9 (2009–2019).
Notes: The 'Separated' category includes separated as well as widowed individuals. The 'Single' category refers to never partnered individuals. Those who become single following union dissolution are included in the 'Separated' category. Those who repartner following separation or widowhood are included in the cohabiting or married category depending on the type of their partnership. PT denotes part-time and FT denotes full-time.

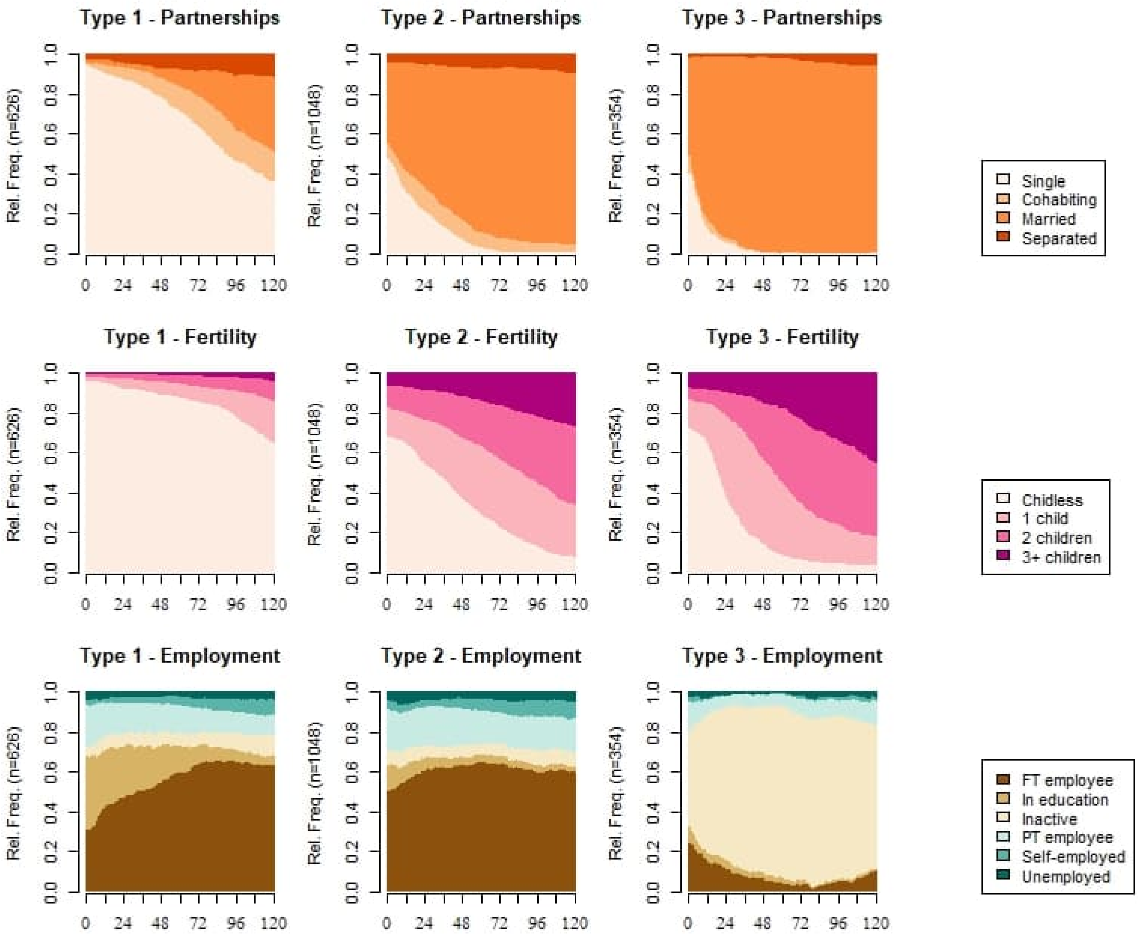
Figure S3: Results of cluster analysis using a three-year observation window (n = 2,724)



Source: Authors' calculations using UKHLS data, waves 1–9 (2009–2019).

Notes: The 'Separated' category includes separated as well as widowed individuals. The 'Single' category refers to never partnered individuals. Those who become single following union dissolution are included in the 'Separated' category. Those who repartner following separation or widowhood are included in the cohabiting or married category depending on the type of their partnership. PT denotes part-time and FT denotes full-time.

Figure S4: Results of cluster analysis using a ten-year observation window (n = 2,028)



Source: Authors' calculations using UKHLS data, waves 1–9 (2009–2019).
Notes: The 'Separated' category includes separated as well as widowed individuals. The 'Single' category refers to never partnered individuals. Those who become single following union dissolution are included in the 'Separated' category. Those who repartner following separation or widowhood are included in the cohabiting or married category depending on the type of their partnership. PT denotes part-time and FT denotes full-time.

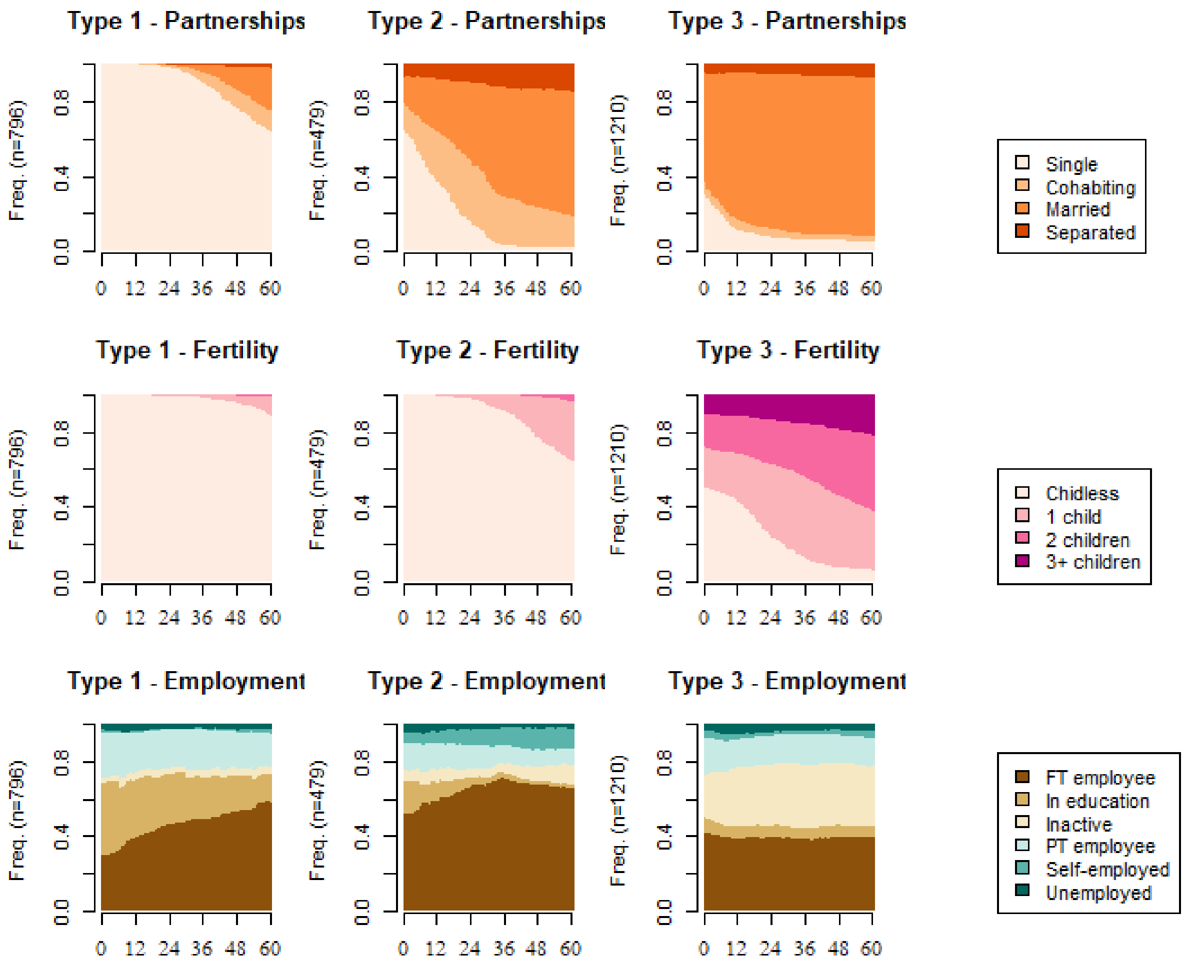
1.3 Choice of the cost regime, distance measure, and clustering method

When applying sequence analysis, researchers need to make several *a priori* decisions, such as the choice of the cost regime (i.e., setting substitution, insertion, and deletion costs), the method to calculate the dissimilarity between pairs of sequences, and the method of cluster analysis. The results can be highly sensitive to these decisions (Brzinsky-Fay and Kohler 2010). Therefore, we have tested the robustness of our results to a range of different specifications.

First, we experiment with using different cost regimes. We have kept the same substitution matrix as in the main analysis and varied the indel costs to be 1.5 and 2. Then, we have used the same indel costs as in the main analyses and changed the substitution matrix to a constant value of 1, 1.5, and 2. Figure S5 shows the results for using a substitution matrix based on the transition rates in the data and indel costs of 1.5. The remaining results are very similar (not shown but available on request). Next, we explore different methods to calculate the distances between pairs of sequences. Keeping the same cost regime as in the main analysis, we have tested the sensitivity of our results using the longest common subsequence (Elzinga and Studer 2015) (not shown but available on request) and the dynamic Hamming distance (Lesnard 2006, 2004) (Figure S6).

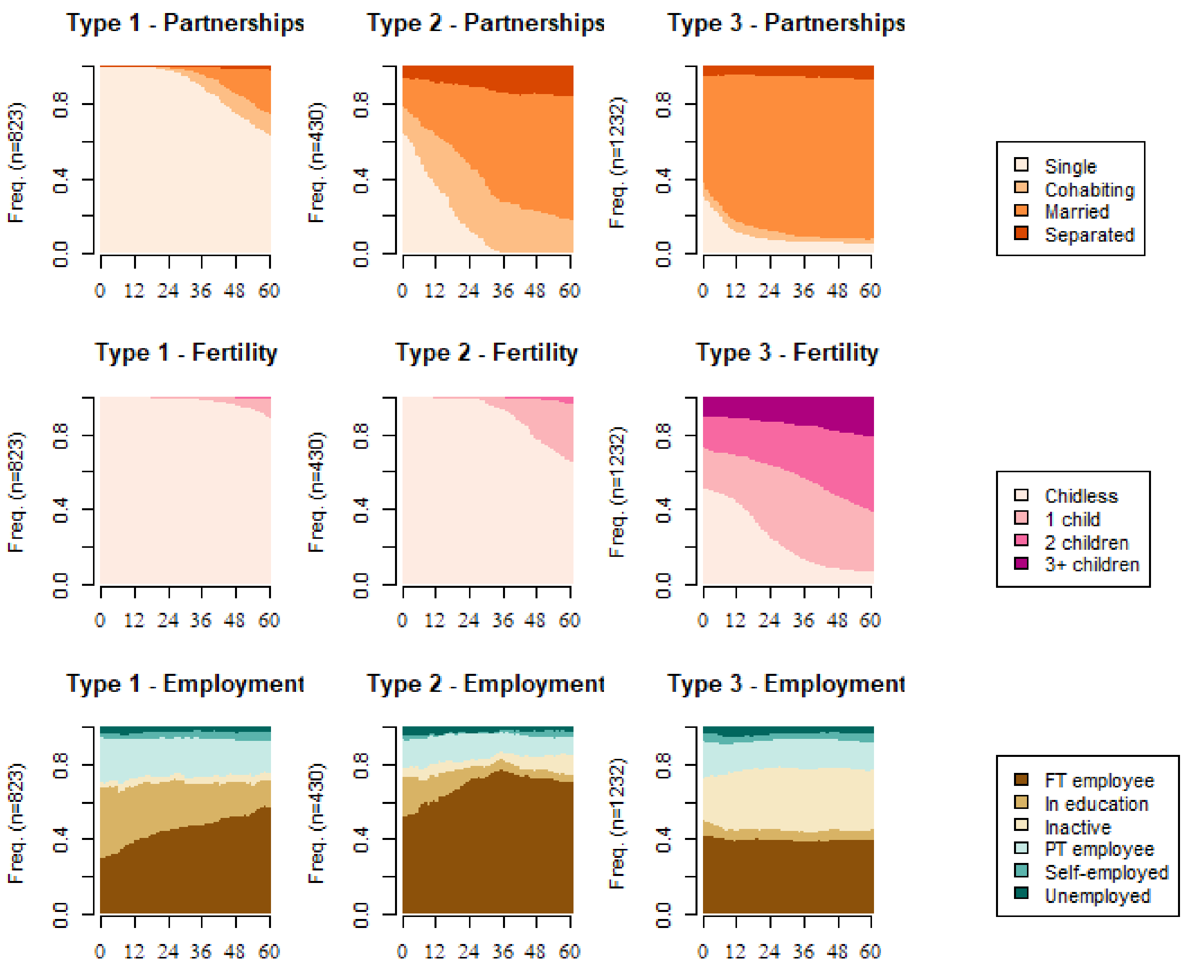
The choice of the cost regime and distance measures only minimally influence the emerging clusters. The largest difference is the way in which individuals are classified into clusters based on their employment histories. In the main analysis, the first cluster is smaller and more homogeneous; it represents individuals who were students or part-time employed at the time of arrival with many transitioning to full-time employment later. In the sensitivity analyses, this cluster is larger and includes many individuals who were full-time employed at the time of arrival. These full-time employed individuals were largely in the second cluster in the main analysis, which is also why we observed a larger proportion of those who arrived as and remained unpartnered in the second cluster. To conclude, these differences are small, mainly pertain to the differential classification of a small number of individuals in the analyses, and do not alter our groupings, the main results, or conclusions. Furthermore, using the specifications selected for the main analysis allows us to better isolate those who arrive to the UK as students.

Figure S5: Results of cluster analysis using a substitution matrix based on transition rates and indel costs of 1.5



Source: Authors' calculations using UKHLS data, waves 1–9 (2009–2019).
Notes: The 'Separated' category includes separated as well as widowed individuals. The 'Single' category refers to never partnered individuals. Those who become single following union dissolution are included in the 'Separated' category. Those who repartner following separation or widowhood are included in the cohabiting or married category depending on the type of their partnership. PT denotes part-time and FT denotes full-time.

Figure S6: Results of cluster analysis using dynamic Hamming distances



Source: Authors' calculations using UKHLS data, waves 1–9 (2009–2019).
Notes: The 'Separated' category includes separated as well as widowed individuals. The 'Single' category refers to never partnered individuals. Those who become single following union dissolution are included in the 'Separated' category. Those who repartner following separation or widowhood are included in the cohabiting or married category depending on the type of their partnership. PT denotes part-time and FT denotes full-time.

Finally, keeping the same costs and distance measures as in the main analysis, we have explored whether and how using different clustering methods influence our results (not shown but available on request). We have used hierarchical clustering with Ward's distance, as well as Partitioning Around Medoids (PAM) using the results of the hierarchical cluster analysis as starting values following Studer (2013). Regardless of the clustering method used, we found the same three clusters as shown in the results section. There are some small differences; while the results of hierarchical cluster analysis are almost identical to those in the main analysis, PAM arrives at slightly different groupings. Cluster 3 is small and mainly consists of inactive and part-time employed individuals. Those who were full-time employed and were classified into cluster 3 in the main analysis are grouped into cluster 1 by PAM. This means that PAM arrives at a much more heterogeneous cluster 1 and a much smaller cluster 3 than the agglomerative nested clustering algorithm used in the main analysis. To conclude, the results are robust to using different cluster analysis methods. Where there are small differences, the agglomerative nested clustering algorithm arrives at more homogeneous and substantively more relevant groupings than the other two methods justifying our choice of using this method.

1.4 Number of clusters

Another challenge during sequence analysis is deciding on the number of clusters. The WeightedCluster package (Studer 2013) includes several measures to assess the quality of the clusters obtained and to help choosing the number of clusters that fits the data best. We have calculated a range of statistics available in the package for a two-, three-, and four-cluster solution (Table S2).

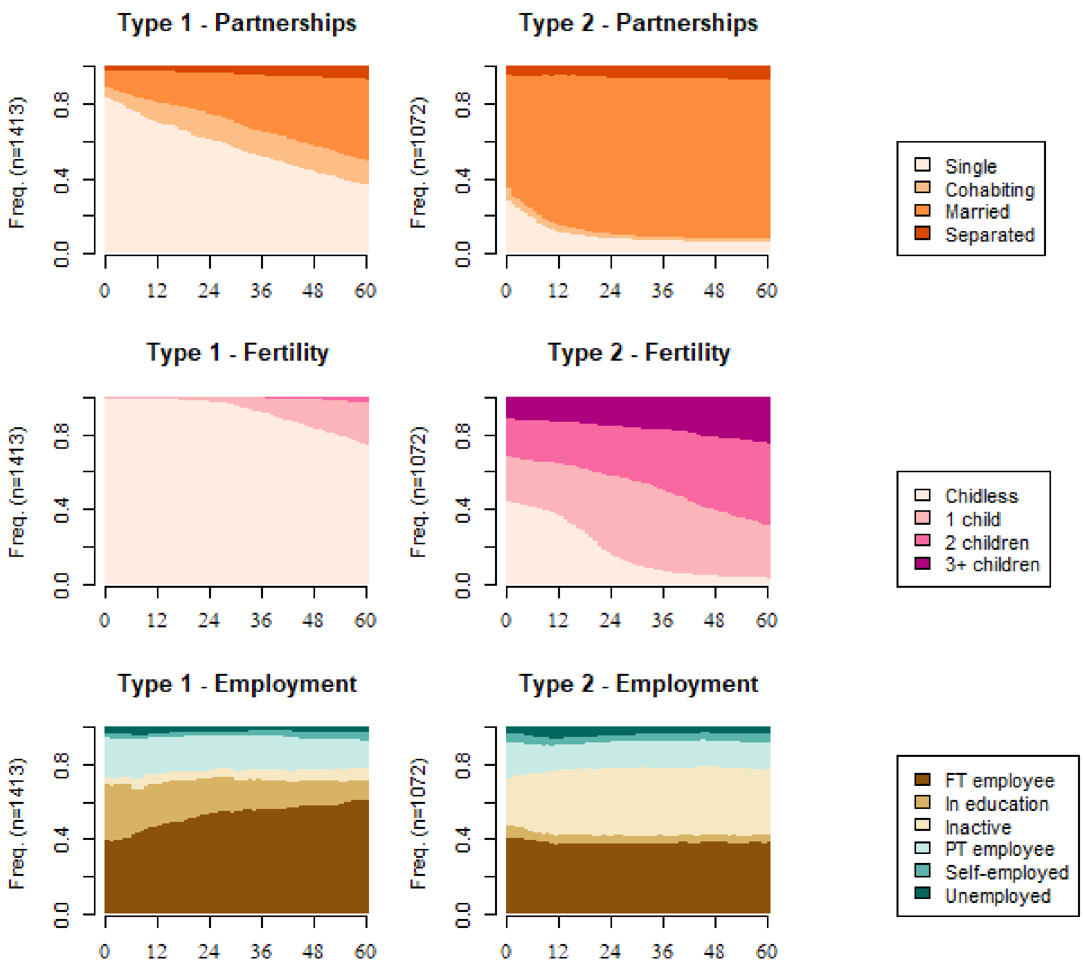
Table S2: Measures of cluster quality for 2, 3, and 4 clusters

	2 clusters	3 clusters	4 clusters
Point Biserial Correlation (PBC)	**0.55**	0.44	0.52
Hubert's Gamma (HG)	**0.64**	0.53	**0.64**
Hubert’s Somers’ D (HGSD)	**0.64**	0.53	**0.64**
Average Silhouette Width (ASW)	**0.32**	0.21	0.23
Average Silhouette Width weighted (ASWw)	**0.32**	0.21	0.23
Calinski-Harabasz index (CH)	**591.16**	431.55	337.62
Pseudo R squared (R2)	0.19	0.26	**0.29**
Calinski-Harabasz index using squared distances (CHsq)	**1225.92**	863.48	721.80
Pseudo R squared using squared distances (R2sq)	0.33	0.41	**0.47**
Hubert's C (HC)	**0.16**	0.23	0.18

Source: Authors’ calculations using UKHLS data, waves 1–9 (2009–2019).
Notes: Boldface highlights the best fit according to each measure. A higher value indicates better fit for all measures except for Hubert’s C, where a lower value indicates better fit.

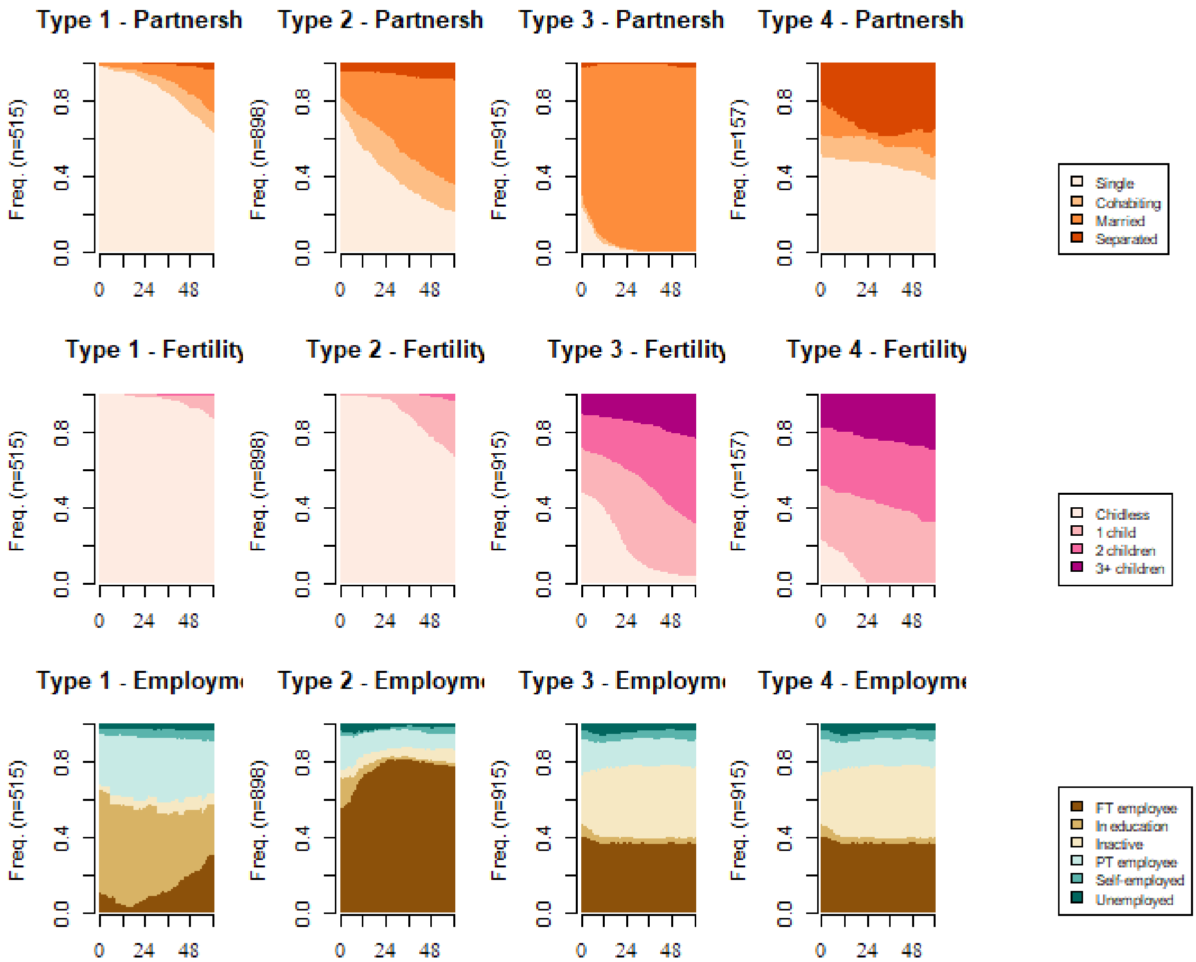
The statistics show inconclusive results; many recommend two or four clusters. Therefore, we have also visually inspected the two-, three- and four-cluster solutions. The two-cluster solution (Figure S7) distinguishes between those who are single, childless, and employed/students and those who are married with children and are either employed or inactive. In the next step of cluster analysis, the first cluster splits into two groups; these are the first two clusters presented in the results section. These two clusters are substantively different, and they still have a large enough number of observations to constitute two separate groups. In the next step of cluster analysis, the third cluster splits into a further two clusters (Figure S8). However, the fourth cluster is very small (157 cases), and these two clusters are very similar regarding fertility and employment trajectories. It contains individuals who were either single or separated. As the fourth group is not large enough and it is not substantively different from the third group, we chose to analyse the three-cluster solution.

Figure S7: Results of cluster analysis: two-cluster solution



Source: Authors' calculations using UKHLS data, waves 1–9 (2009–2019).
Notes: The 'Separated' category includes separated as well as widowed individuals. The 'Single' category refers to never partnered individuals. Those who become single following union dissolution are included in the 'Separated' category. Those who repartner following separation or widowhood are included in the cohabiting or married category depending on the type of their partnership. PT denotes part-time and FT denotes full-time.

Figure S8: Results of cluster analysis: four-cluster solution



Source: Authors' calculations using UKHLS data, waves 1–9 (2009–2019).
Notes: The 'Separated' category includes separated as well as widowed individuals. The 'Single' category refers to never partnered individuals. Those who become single following union dissolution are included in the 'Separated' category. Those who repartner following separation or widowhood are included in the cohabiting or married category depending on the type of their partnership. PT denotes part-time and FT denotes full-time.

1.5 Validity of the MCSA approach

A recent study by Piccarreta (2017) argued that a joint analysis of several domains is only reasonable and effective if these domains are also statistically associated with each other. She proposed several measures to calculate the strength of the interrelationship between more than two domains. We calculated Pearson's correlation and Cronbach's alpha to measure the statistical association between the domains of partnership, fertility, and employment (Table S3). According to these measures, the association between each pair of domains as well as all three domains is weak to moderate. However, we know from the empirical literature that these domains are interrelated in individuals' lives. Therefore, we deem it justified from a substantive point of view to analyse how partnership, fertility, and employment trajectories jointly evolve in the lives of immigrants.

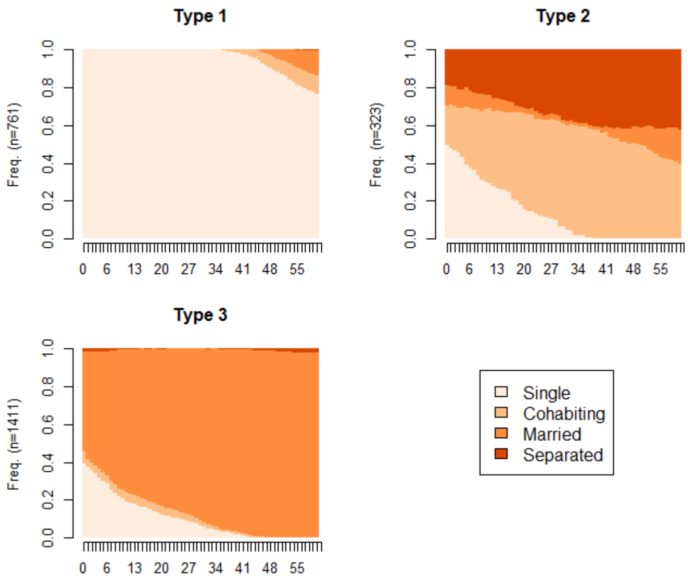
Table S3: Measures of correlation between the three domains of partnership, fertility, and employment

	Pearson's correlation		
	Partnership	Fertility	Employment
Partnership	1		
Fertility	0.161	1	
Employment	0.049	0.057	1
Joint	0.655	0.670	0.544
	Cronbach's alpha		
Partnership & fertility		0.277	
Partnership & employment		0.094	
Fertility & employment		0.107	
All three domains		0.227	

Source: Authors' calculations using UKHLS data, waves 1–9 (2009–2019).

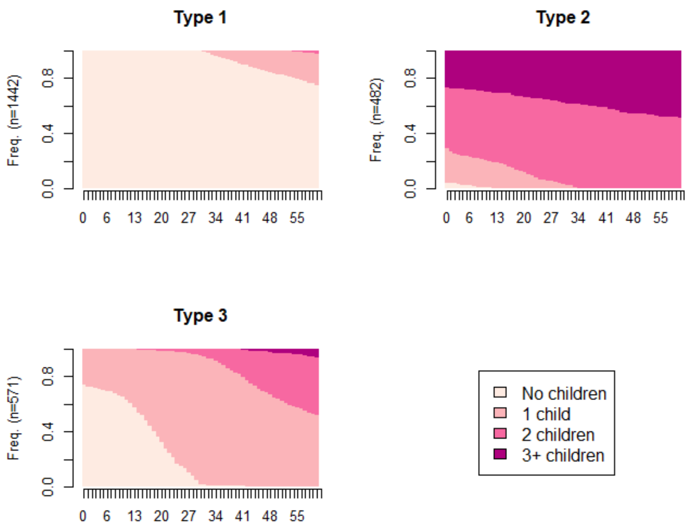
Second, it is recommended to check that the results of MCSA are not dominated by patterns in just one domain (Piccarreta 2017). We have conducted single-channel SA separately for partnership, fertility, and employment trajectories. Regarding partnership trajectories (Figure S9), we found the following three clusters: those who are and remain never partnered, those who form or dissolve partnerships, and those who are married. The fertility clusters are childless, those who arrive with two or more children, and those who have a(nother) child following migration to the UK (Figure S10). We also found three employment clusters: employed, students and part-time employees, and those who are inactive (Figure S11). Overall, these groups are similar to what is identified in the MCSA but the joint analysis of these three domains allows us to highlight how these different domains intersect and jointly evolve in the lives of migrants after their arrival in the UK.

Figure S9: Results of cluster analysis using single channel sequence analysis: Partnership trajectories



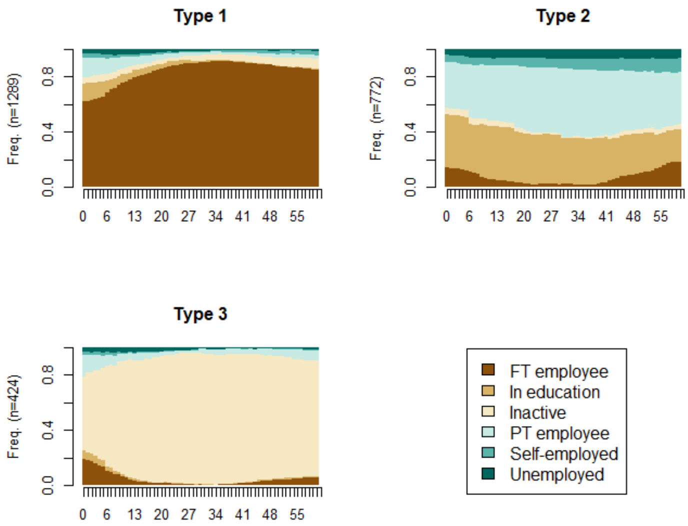
Source: Authors' calculations using UKHLS data, waves 1–9 (2009–2019).
Notes: The 'Separated' category includes separated as well as widowed individuals. The 'Single' category refers to never partnered individuals. Those who become single following union dissolution are included in the 'Separated' category. Those who repartner following separation or widowhood are included in the cohabiting or married category depending on the type of their partnership.

Figure S10: Results of cluster analysis using single channel sequence analysis: Fertility trajectories



Source: Authors' calculations using UKHLS data, waves 1–9 (2009–2019).

Figure S11: Results of cluster analysis using single channel sequence analysis: Employment trajectories



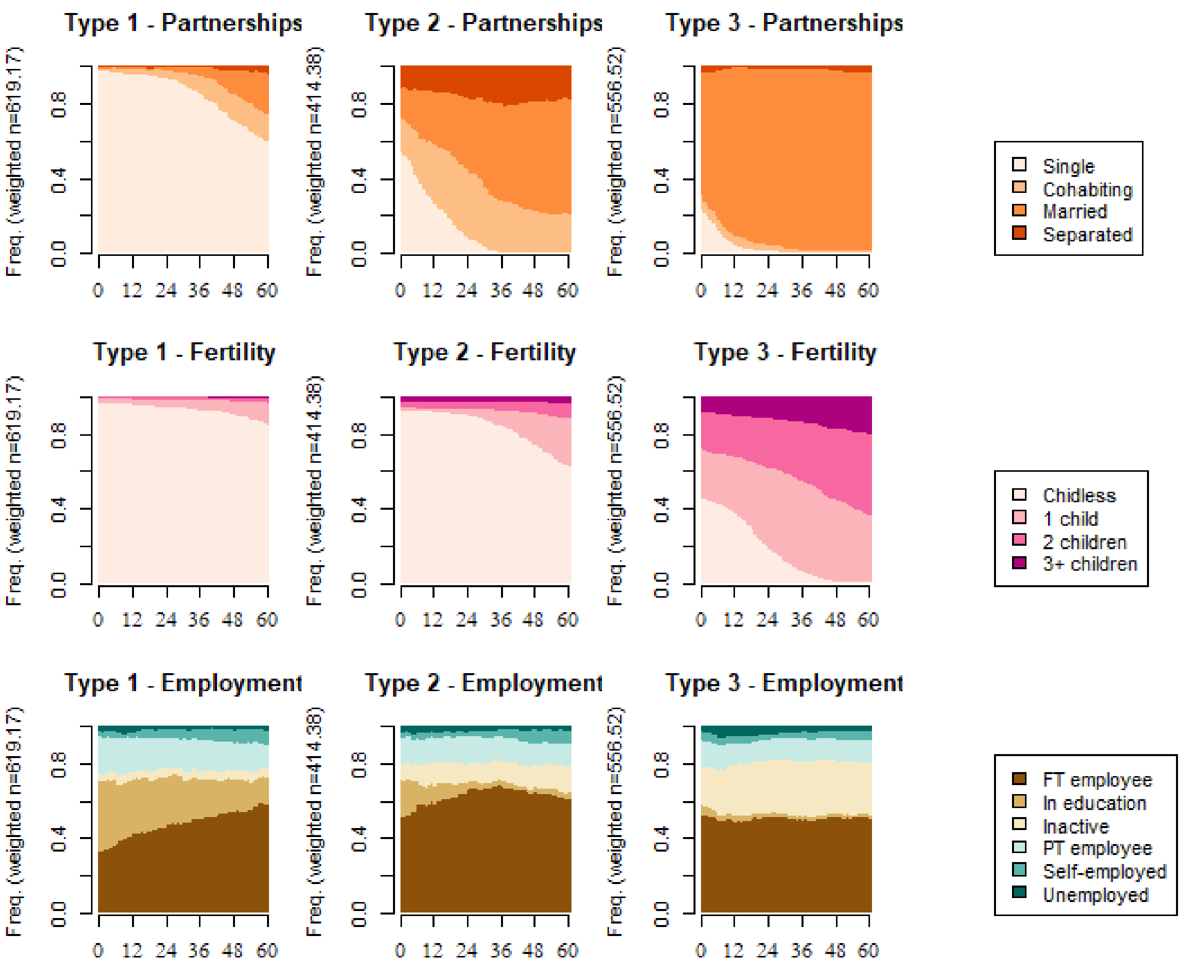
Source: Authors' calculations using UKHLS data, waves 1–9 (2009–2019).
Notes: PT denotes part-time and FT denotes full-time.

1.6 Weighted analysis

The sampling design of the UKHLS is complex and the immigrant and ethnic minority boost samples were taken from high ethnic minority concentration areas. Therefore, it is argued that to provide unbiased population estimates, researchers need to use weights and to account for the complex sampling strategy (McFall, Nandi, and Platt 2019). In the results section we have presented unweighted results for two reasons. First, we only focus on immigrants and do not compare them with the native British. This means that even if immigrants are overrepresented in the data in comparison to natives, this issue is not relevant for our analysis. Second, it is not trivial which weights researchers should use when retrospective information is utilised, and immigrants enter the country and the dataset at different time points.

Nonetheless, we have conducted robustness checks using cross-sectional weights from the first wave (2009/10) of UKHLS. 135 individuals were excluded from the analysis because they were not yet included in the dataset at wave 1 meaning that they do not have a value for the cross-sectional weight. Thus, we conduct the weighted analysis on a sample of 2,350 individuals. We incorporate weights both into the definition of the sequence objects and during clustering. The weighted results (Figure S12) are very similar to what is shown in the main analysis; the largest difference is again in cluster 1 regarding employment histories; rather than isolating students, this group is a mix of students and full-time employed individuals.

Figure S12: Results of weighted cluster analysis (n = 2,350)



Source: Authors' calculations using UKHLS data, waves 1–9 (2009–2019).
Notes: The 'Separated' category includes separated as well as widowed individuals. The 'Single' category refers to never partnered individuals. Those who become single following union dissolution are included in the 'Separated' category. Those who repartner following separation or widowhood are included in the cohabiting or married category depending on the type of their partnership. PT denotes part-time and FT denotes full-time.