Dynamic Analysis of Work-Life Histories: Manual

This manual explains and advises on use of code to analyse dynamic features of work-life histories contained in the LAMAFLO GitHub.

What dynamic analysis does the code enable?

Data organised by pidp-Spell

- Creates a transition variable (capturing current and previous state) and date of transition (current spell start), at each spell start date. ("Transition" consists of remaining in the same state if there is no change in status.)
- Spell durations of each spell. (Spell duration is measured at the end of each spell, so the relevant date is the spell end date.)
- Durations to date in each status type. (Durations are cumulatively measured up the end of each spell, so the relevant "to date" is the spell end date.)
- Counts over work-life history to date of spells in user-chosen statuses. The counts include the current spell, so relate to spells that have all been completed at each spell end date. Between the spell start and spell end date, the count is correct but includes the current incomplete spell.
- Counts over work-life history to date of transitions, for each transition type. Th is cumulative measure of transitions is valid at the current spell start date.
- Conversion of data organised by pidp-spell to data organised by pidp-month. This monthly data is available in either 'wide' or 'long' form. The code as written creates labour force status at monthly frequency, for the user-specified states and over a user-specified time interval (using optional age restrictions). The code could be adapted to collect other variables in monthly form.
- Creation of labour market flows and flow rates based on user-specified states.

Data organised by pidp-Wave (Interview Date)

Research may require work-life history information to be used alongside more typically-organised data for individuals obtained from each Wave. Code is provided to obtain potentially useful variables capturing work-life dynamics that can be merged directly with Wave-based data, as they are organised by pidp-Wave. Because participant responses are obtained at interview date, and often relate to that time (e.g. wage and other job-related information; satisfaction and opinions; location), summarising individuals' experience of dynamic work-life events up to the interview date is likely to be relevant.

- Time spent in each of the user-chosen states, over the individual's work-life history, up to the interview date.
- Duration of the current spell at interview date.
- Start and end dates of the spell held at interview date (the end date being after the interview date, of course – unless, according to the individual's work-life history, the spell ends at interview date – apart from the last interview date in the sample, where the end date is recorded as the interview date).
- Count of spells in each of the user-chosen statuses, over work-life history, up to interview date.
- Count of transitions, by transition type, over work-life history, up to interview date.
- Also provided are the individual's first and last interview waves and interview dates.

Defining states

The user is asked to specify the states they wish to analyse and the definitions of those states in terms of the underlying values of the original `raw' data's status variable.

The 'raw' data used for the dynamic analysis must contain a variable describing the status (state, activity) of the individual. In LW-based data, this variable is named `Status`. For default settings in the updated "`Launch programs.do`", `Status` takes the values shown in the second column of Table 1 (values taken from the `jbstat` variable in BHPS and UKHLS data). Column 3 shows a possible classification for dynamic analysis. Dynamic analysis is generally best with a small well-defined status, and column 3 uses a conventional classification of categories into three states (employment, unemployment, nonparticipation).

Table 1: Tabulation of Status variable from Merged Dataset (using default settings) with an example of a possible definition of states for dynamic analysis

Economic activity	Code	Classification for 3-state dynamic analysis	Freq.	Percent
Self-Employed	1	E	30,758	5.88
Paid Employment	2	E	240,992	46.05
Unemployed	3	U	40,789	7.79
Retired	4	N	34,343	6.56
Maternity Leave	5	E	8,942	1.71
Family Carer or Homemaker	6	N	34,529	6.60
Full-Time Student	7	N	104,858	20.04
Sick or Disabled	8	N	12,358	2.36
Govt Training Scheme	9	E	3,370	0.64
Unpaid, Family Business	10	E	216	0.04
Apprenticeship	11	E	378	0.07
Doing Something Else	97	N	9,318	1.78
Paid Work	100	E	774	0.15
Something Else	101	N	24	0.00
National Service/War Service	103	E	1,686	0.32
Total			523,335	100.00

Input data required

The code to generate dynamic variables uses `raw' data initially organised by pidp-Spell.

Merged Dataset.dta

Any final dataset resulting from analysis based on code by Liam Wright (2020) (LW) would be suitable. That code, run via "`Launch Programmes.do`", produces a final dataset "`Merged Dataset.dta`".

"`Merged Dataset.dta`" is the default name used for the data file to be analysed by "`Dynamics - Launch Program.do`", the overarching dynamic analysis .do file.

Interview Grid.dta

To produce datasets organised by interview date (an option that users can turn off or on within "`Dynamics - Launch Program.do`"), information in the dataset "`Interview Grid.dta`" is required in addition to "`Merged Dataset.dta`" (or a similarly-structured dataset). "`Interview Grid.dta`" is also produced by "`Launch Programmes.do`".

Outputs

- "Dynamics_[name of user input data file].dta" (which could be "Dynamics_Merged Dataset.dta" if the input data results from analysis based on LW code) will contain transitions, durations and counts.
- Users can select to output work-life history information about (user-chosen) statuses for each individual for every month, in wide or long form. Users can set the date at which they wish calculation of flows to begin, and monthly data on status will be produced from that date onwards. Optional outputs: "Dynamics_lfs_wide_pidp_m.dta", "Dynamics_lfs_long_pidp_m.dta". Additional optional outputs if weighted flows are selected: "Dynamics_weight_wide_pidp_m.dta", "Dynamics_weight_long_pidp_m.dta".
- If the relevant options are selected, "Dynamics_Flows_M/Q/Y[_W].dta" will contain flows and flow rate estimates, with separate files output for weighted (_W) and unweighted flows and separate files for each time dimension (monthly/quarterly/annual).
- If the option is selected to generate data by interview date, a file "Dynamics_All_AtInt.dta" is output containing all of the following:
  - Time spent in each (user-chosen) status for each individual up to each interview date.
  - Duration of the spell current at interview, for each individual. Relatedly, start and end dates for the spell current at interview (for the latest interview date, the spell end date will be the interview date).
  - Counts of spells in each (user-chosen) status up to and including the interview date.
  - Counts of transitions between (user-chosen) statuses up to and including the transition leading to the status held at interview.
  - This file also contains information for each individual on first and latest interview dates and first and latest waves in which the interview occurred.

User choices in "Launch Program_Dynamics.do"

- Users need to specify the names of the variables in their `raw' dataset that measure:
  - Status default name: Status
  - Spell start date Start_MY
  - Spell end date End_MY
  - Interview date IntDate_MY

  Birth date is required if you wish to restrict the sample by age (default variable name: Birth_MY); if this variable is not available, no age restriction will be imposed.

  The default names correspond to the variable names used in dataset "Merged Dataset.dta" produced by "Launch Programmes.do".

- Users need to define the states they wish to analyse dynamically:
  - Definitions in terms of values of the `raw' data status variable (see Table 1 for an example).
  - Names for the states to be analysed (e.g. "E", "U", "N"). These names should be single letters.
- Users can choose to restrict analysis to a particular age range by setting minimum and maximum ages.
- Users wanting to create flows can choose:
  - The date at which they want flows data to start.
  - Whether monthly, quarterly, or annual flows are required (any combination can be chosen).
  - Unweighted flows (counts) or weighted flows (or both).

Choices for flows: Advice

- Users can choose quarterly, annual or monthly flows. The choice should depend on user needs, but should take into account that the number of observations per state decreases with frequency of observation and number of states, so flows are less volatile if they are based on a lower-frequency time dimension and more parsimonious states.
- The default choice of date at which flows calculation will start is `1990jan`. Choosing January 1990 will output flows and flow rates for the full BHPS-UKHLS survey period but exclude earlier data based only on recall. The choice of date at which flow calculation starts will depend on research needs. The longer the sample, the more time flows calculation takes – but, importantly, the relative lack of data before the BHPS survey started (when work-life histories are based on life history recall) means calculation over those years is relatively rapid. The start of the BHPS survey and the end of the UKHLS survey are also characterised by relatively few observations and quick flows calculations.
- Flow rate calculations are sensitive to cell size. The lack of data pre-BHPS survey start and at the start of BHPS and end of UKHLS samples makes flow rates volatile. Users may wish to inspect the estimates and trim the sample to exclude those periods if required.
- Users can choose unweighted or weighted flows. Unweighted flows are simple counts of transitions between states. Weighted flows are intended to give estimates representative of the British population (BHPS and UKHLS sample).
    - *Choice of weights for flows.* Flows capture the change in state between time periods where the data for those time periods often does not arise from observation at that time period. Some states are observed contemporaneously, but many work-life-history states are the result of later reports (recall). Although flows are clearly a longitudinal measure, the genesis of the data means that provided longitudinal weights are not necessarily any better than cross-sectional weights. Longitudinal weights relate to Wave-to-Wave changes; life-history records generally do not match that time frame. So longitudinal weights are not necessarily appropriate for life-histories. Cross-sectional weights are available for more of the sample which means that weighted estimates using these weights retain more valuable and relatively scarce life-history information. Use of cross-section weights in flows creation has precedence (e.g. Smith, 2011). For weighted flows, cross-section weights are selected that relate to the interview date at which spell and status information was obtained.

Time taken for the analysis

The dynamic analyses to calculate transitions, durations and counts are fast.

The optional creation of flows is relatively time-consuming. The conversion to pidp-month data is slow as information has to be allocated to each month for each individual. For example, taking individuals in the "`Merged Dataset.dta`" (default settings) over the 386 months from January 1990 to March 2022, there are over 31 million observations. The length of time to create flows will depend on set-up, but it is normal for weighted flows and flow rate calculation to take around 1.5 hours. Weighted flows take slighly longer than unweighted (due to the need to assign weights to each individual for each month).

The generation of data by interview date (wave) takes some time as the code has to search across spell data to allocate the correct information to each interview date. Generation of all interview-date dynamic variables collected in "`Dynamics_All_AtInt.dta`" may take in the region of 45 minutes.