Documento de Trabajo - 2017/06

Faster estimation of discrete time duration models with unobserved heterogeneity using hshaz2

David Troncoso Ponce
(Universidad Pablo de Olavide)

fedea

Faster estimation of discrete time duration models with unobserved heterogeneity using hshaz2

David Troncoso Ponce

Universidad Pablo de Olavide

RESUMEN

En este trabajo se presenta el nuevo comando de estimación **hshaz2** y se describen sus principales características. Al igual que el anterior comando **hshaz**, desarrollado por el profesor Stephen Jenkins, el comando **hshaz2** estima, mediante el método de máxima verosimilitud, un conjunto de modelos de duración en tiempo discreto con riesgos proporcionales que tiene en cuenta la presencia de heterogeneidad inobservable. La aportación principal del comando **hshaz2** consiste en el desarrollo y programación de las expresiones algebraicas de las primeras y segundas derivadas parciales de la función de log-verosimilitud, que componen el vector gradiente y la matrix Hessiana, respectivamente.

De este modo, la diferencia entre **hshaz** y **hashaz2** reside en el método empleado para alcanzar la convergencia de la función de log-verosimilitud: el comando **hshaz** usa el método d0, mediante el cual las derivadas del gradiente y el Hessiano son calculadas usando aproximaciones numéricas; mientras que **hshaz2** usa el método d2, que provee las expresiones algebraicas de las derivadas de primer y segundo orden.

Las ventajas de emplear el nuevo comando **hshaz2** son fundamentalmente dos: la mayor fiabilidad de los errores estándar de los parámetros estimados; y sobre todo, la reducción de los tiempos de computación requeridos para la estimación de este tipo de modelos, que posibilita la estimación de modelos de duración empleando grandes bases de microdatos longitudinales.

Faster estimation of discrete time duration models with unobserved heterogeneity using hshaz2

David Troncoso Ponce
Pablo de Olavide University
Seville, Spain
Email: dtropon@upo.es

Abstract. This article presents `hshaz2`, a new Stata command that uses `d2 ml` method to estimate discrete time duration models with unobserved heterogeneity. The main advantage of using `hshaz2` is the gain in computation speed, that takes special relevance as the sample size increases. Estimation results show that, on a sample size of 568,042 observations, `hshaz2` spends 0.42 and 1.13 minutes to achieve the convergence of a discrete time proportional hazard model with two and three points of support, respectively. Furthermore, `hshaz2` allows for the estimation of multispell duration models, where individuals may be observed at risk of exiting more than once. Using, a sample with 1,547,507 observations, `hshaz2` spends 1.17 and 2.17 minutes to achieve convergence of a multi-spell discrete time proportional hazard model with two and three points of support, respectively.

Keywords: Duration analysis, Unobserved heterogeneity, `d2 ml` method, `hshaz`, `hshaz2`, Hessian matrix

Acknowledgement

I am greatful to professor Stephen Jenkins for his helpful comments and suggestions that have contributed to significantly improve this work, and for allowing me to use Stata code of his hshaz's command syntax and helpfile. I also thank financial support of research project SEJ-6882 from Junta de Andalucía. This article has been finished during a postdoctoral research stay at Fundación de Estudios de Economía Aplicada (FEDEA).

1 Introduction

Time required to estimate discrete time duration models that account for the presence of unobserved heterogeneity (hereafter, UH) uses to be an important concern for applied researchers when have to deal with large datasets.[1] Stata command `hshaz`, written by professor Stephen Jenkins, estimates proportional discrete time duration models taking into account the the presence of UH, following to [Heckman and Singer, 1984]. `hshaz` uses `d0 ml` method to achieve

[1] For the purpose of this article, I consider a large dataset as those with at least one million observations, as for example, longitudinal data that comes from Social Security administrative records.

convergence of the log-likelihood function, which computes numerical approximation to first and second order derivatives, that composes gradient vector and Hessian matrix, respectively.

This article presents `hshaz2`, a new Stata command that provides the algebraic expressions of both first and second order partial derivatives of the log-likelihood function estimated by `hshaz` command, and describes its main characteristics. Thus, `hshaz2` also estimates, using maximum likelihood method, discrete time proportional hazard rate models with UH. However, `hshaz` and `hshaz2` differ in the Stata `ml` method used to achieve convergence of the log-likelihood: `hshaz` uses `d0 ml` method, whereas `hshaz2` uses `d2 ml` method.

The main advantage of programming the algebraic expressions of first and, above all, second order derivatives of the log-likelihood function are twofold: First, the reliability on the standard errors of parameters estimates. Second, and the most important issue, the gains obtained in computation speed to achieve the model convergence using `d2` method [Gould et al., 2010].[2]

The rest of the article estructures as follows: Section 2 describes the database used to obtain estimation results; the econometric model and `hshaz2` command syntax are explained in Sections 3 and 4, respectively; Section 5 presents estimation results, and Section 6 describes some details on the reparameterization of mass-points probability parameters. Finally, Section 7 concludes.

2 Database: The Continuous Sample of Working Histories

I analyze a longitudinal sample composed of 44,077 unemployed workers, aged 16-37 year-old, in the Spanish labor market for the period 2000-2013, that comes from the *Continuous Sample of Working Histories* database (CSWH, hereafter). The CSWH is a longitudinal database that provides the working histories records of more than one million people, who represent a 4% non-stratified random draw from a target population, composed of any person with a contribution relation with the Spanish Social Security Administration. It includes both wage workers and recipients of Social Security benefits, namely, unemployment benefits, disability, survivor pension and maternity leave.[3]

The CSWH contains detailed information on each employment and unemployment episodes experienced by workers through their entire working histories. The information provided by the CSWH can be grouped into several categories: First, personal characteristics of workers (gender, age, nationality, educational

[2] A detailed explanation, using an applied approach, on the estimation of duration models using `hshaz` command is available at the following link: http://www.iser.essex.ac.uk/teaching/stephenj/ec968/index.php

[3] [García-Pérez, 2008] and [Lapuerta, 2010] contain a deep exposition about features of CSWH as well as all necessary techniques to perform a duration analysis using working lives information.

level, residence place, etc); Second, job characteristics (type of labor contract, part-time coefficient, qualification level, etc); Third, information on the employer (firm size, activity sector, etc). Furthermore, an important feature of the CSWH is that provides the beginning and termination dates of all employment and unemployment episodes, which takes special interest for duration analysis.

For the unispell duration model estimation, that will be explained in Subsection 5.1, each unemployment episode has been expanded in monthly intervals. Thus, each unemployed worker is observed (and, therefore, contributes to the likelihood function) as many times as the number of months the unemployment episode lasts. After the database has been expanded, the sample size increases up to 568,042 observations. Moreover, the duration variables as well as all explanatory variables that vary with unemployment duration (such as, age, squared age, etc) have been generated in order to correctly measure time-varying covariates.

3 Econometric model

This Section briefly describes the main features of the econometric models that will be estimate in 5. The main goal of this kind of models is to analyze duration spent by a population in a specific state (in this example, unemployment state), as well as to analyze the set of factors, observable and above all unobservable, that affect time spent in that state.

Let's consider an individual beginning an unemployment episode at time $T = 1$ (time T is measured in month intervals). The unemployed worker is observed monthly during the unemployment episode until either he/she finds a new job, or the observation window ends. Unemployment duration is analyzed by estimating the hazard rate out of unemployment at each observed month.

The hazard rate out of unemployment estimated by `hshaz2` (and `hshaz`) takes the following functional form:

$$h(t|x, \eta) = 1 - exp(-exp(\lambda(t) + x\beta + \eta)) \quad (1)$$

As the expression above shows, the hazard rate at month $T = t$ depends on time (months) spent in the current unemployment state (i.e. duration dependence), captured by $\lambda(t)$, as well as on a set of covariates summarized by x vector.[4] Furthermore, the hazard rate also depends on an unobserved component given by η, that mesasures factors, such as job search effort, motivation, ability, etc, that are unobserved to the researcher and may affect the transition rate out of unemployment.

To estimate the unobserved heterogeneity distribution, following [Heckman and Singer, 1984], it is assumed the existence of different types of unemployed workers who differ between them in unobserved characteristics (such as, as mentioned above,

[4] x vector may contain both time-fixed and time-varying covariates.

motivation, ability, etc), that affect the transition rate out of unemployment. Therefore, the whole population is composed of a discrete mixture distribution of all the types of unemployed workers considered by the econometric model. The presence of each type of unemployed workers in the whole population is weighted by the probability of observing it, that is estimated jointly with the rest of the model parameters.

The contribution to the likelihood function of an individual i is given by the following expression:

$$L_i = \sum_{j=1}^{P} \pi_j \{ \prod_{t=1}^{T_i} \frac{h(T=t|\lambda(t), x_{it}, \eta_j)^{y_{it}}}{(1-h(T=t|\lambda(t), x_{it}, \eta_j))^{(1-y_{it})}} S(T=t|\lambda(t), x_{it}, \eta_j)^{(1-y_{it})} \} \tag{2}$$

Where $h(T=t|\lambda(t), x_{it}, \eta_j)$ and $S(T=t|\lambda(t), x_{it}, \eta_j)$ denote the hazard rate and the survival function[5] observed at month $T=t$, respectively, conditional on the duration dependence $\lambda(t)$, on the set of covariates x_{it}, and on belonging to the type of unemployed workers with unobserved characteristics given by η_j.[6]

The discrete probability distribution of unobserved heterogeneity is given by the estimation of the vector $(\pi_1, \pi_2, ..., \pi_P)$, with $\pi_1 = 1 - \sum_{l=2}^{P} pi_l$ and $\pi_j = \frac{e^{p_j}}{1+\sum_{l=2}^{P} e^{p_l}}$, $j = 2, 3, ?, P$. Each π_j parameter estimates the probability of observing each type of (unemployed) workers (in) of the whole population.

Dependent variable $y_{it} = 0, 1$ denotes a dummy variable that takes value 1 if worker i exits out of unemployment at month $T=t$, and takes value zero otherwise.[7]

Finally, the total likelihood function is given by:

$$L = \sum_{i=1}^{N} \sum_{j=1}^{P} \pi_j \{ \prod_{t=1}^{T_i} \frac{h(T=t|\lambda(t), x_{it}, \eta_j)^{y_{it}}}{(1-h(T=t|\lambda(t), x_{it}, \eta_j))^{(1-y_{it})}} S(T=t|\lambda(t), x_{it}, \eta_j)^{(1-y_{it})} \} \tag{3}$$

`hshaz` command maximizes, using `d0 ml` method, the natural logarithm of L to estimate (the all) model parameters. The main contribution of `hshaz2` command is that it provides the algrebraic expressions of both the first and second order (partial) derivatives of the natural logarithm of L, and therefore, achieves the convergence using `d2 ml` method.

4 Command syntax

As has been explained in Section 1, the main contribution of `hshaz2` command is the programming of both first and second order derivatives to achieve faster

[5] $S(T=t|\lambda(t), x_{it}, \eta_j) = ((1-h(T=1|\lambda(1), x_{i1}, \eta_j)))((1-h(T=2|\lambda(2), x_{i2}, \eta_j)))...((1-h(T=t-1|\lambda(t-1), x_{it-1}, \eta_j)))((1-h(T=t|\lambda(t), x_{it}, \eta_j)))$

[6] It is assumed that unobserved characteristics do not vary with time and are not correlated to the rest of explanatory variables included in the specification of the hazard rate.

[7] Dependent variable y_{it} refers to `dead(deadvar)` of `hshaz2` command.

estimations by using `d2 ml` method. The programming of gradient vector and Hessian matrix do not affect the estimation output reported by the former **hshaz** command, in the sense that the set of parameters to be estimated is the same, either by using `d0` or `d2 ml` methods, respectively. Therefore, the command syntax of **hshaz2** shares the same estructure that **hshaz**'s, and also makes use of the same terminology to refer to the estimation output. This section describes the main estructure of **hshaz2** command syntax and explains the options allowed by **hshaz2** for the estimation process.

The **hshaz2** command syntax is:

`hshaz2` *varlist* [*weight*] [*if exp*] [*in range*] [, `id(`*idvar*`)` `dead(`*deadvar*`)` `seq(`*seqvar*`)` `spell(`*spellvar*`)` `nmp(`#`)` `m2(`#`)` `p2(`#`)` `m3(`#`)` `p3(`#`)` `m4(`#`)` `p4(`#`)` `m5(`#`)` `p5(`#`)` `eform` `nocons` `nolog` `nobeta0` `level(`#`)` `maximize_options`]

As can be seen, the only element added by **hshaz2** to the (**hshaz**'s) command syntax is an `option(.)`, called `spell(`*spellvar*`)`. The **hshaz2** command allows for the estimation of multiple spells duration models, by which individuals may be observed at risk of exiting more than once. In such cases, it is necessary to correctly identify and to sort the different episodes experienced by each person in the estimation sample. This is the goal of the option `spell(`*spellvar*`)`, where *spellvar* must be a numeric variable that identifies the sequential order of the spells experienced by each individual in the sample. This will be explained in detail in Subsection 5.2.

5 Estimation

This Section presents results of fitting discrete time duration models on the sample of unemployed workers described in Section 2. The aim of this Section is to show time saving involved by using **hshaz2** command in comparison with **hshaz**, highlighting the importance of using `d2 ml` method to achieve convergence. In 5.1, I use both **hshaz2** and **hshaz** commands to estimate two unispell duration models, with two and three mass-points, respectively. Once the results are obtained, the estimation speed of **hshaz2** and **hshaz** commands are compared. Finally, in Subsection 5.2, I use **hshaz2** command to estimate two multispell duration models, with two and three points of support for the identification of the unobserved heterogeneity, respectively.[8]

5.1 Fitting unispell duration models using hshaz and hshaz2

Pages 7 and 8 present the estimation output of fitting a duration model with two mass-points, running **hshaz2** and **hshaz** commands, respectively. The rest

[8] I work with Stata 14.0 MP - Parallel edition 64 bits. The machine employed to obtain estimation results incorporates an Intel(R) Core(TM) i7-6700HQ CPU at 2.60 GHz, and 12 Gb RAM memory. The operating system is Windows 10 Home, and Stata 14 MP version.

of tables with detailed estimation results are shown at final Appendix. Comments on coeficients estimates will only focus mainly on the different duration dependence effect shown by estimation output with and without controlling for the presence of unobserved heterogeneity. The interpretacion of the rest of estimated coefficients are not commented, because of the main purpose of these regressions is to highlight that `hshaz2` command replicates the results obtained by `hshaz` command, and therefore, it is not intended to address a regression analysis to properly estimate the effect of a set of covariates on the probability of exiting out of unemployment.

The set of covariates is included in the specification of the hazard rate for control purposes, and summarizes: 1) personal characteristics of the unemployed workers, such as, gender, age and squared age,[9] nationality,[10] and educational level; 2) characteristics of the current unemployment spell, such as, a dummy variable to identify whether the unemployed worker receives unemployment benefits, as well as an interaction between this dummy variable and the natural logarithm of the duration of current unemployment spell; 3) the quarterly unemployment rate to capture bussiness cycle effects on the transition rate out of unemployment; 4) a set of dummy variables that identify the Spanish regions to capture regional effects. Additionally to the duration dependence specification (using a two order polinomial of the natural logarithm of the duration of current unemployment spell), three dummy variables are included to identify months 12, 18 and 24. These dummy variables are included to capture exit peaks, frequently observed in unemployment duration analysis, that may be due to unemployment benefits exhaustion effects.

As most empirical research on Labor Economics has found for many European labor markets, coefficient estimates show negative unemployment duration dependence, and reveal the importance of controlling for the presence of UH. Thus, the two-mass points duration model that does not control for the presence of UH [11] underestimates (1.430829 and -0.5548032, for $Log(t)$ and $Log(t)^2$, respectively[12]) the effect of unemployment duration dependence. Controlling for UH, the effect of duration dependence increases up to 3.614478 and -0.9491475, for $Log(t)$ and $Log(t)^2$, respectively.

Regarding the estimation of the unobserved heterogeneity distribution, 71.6% (0.7163904) of the sample are Type I unemployed workers, and the rest 28.4% (0.2836096) belongs to Type II group. For the correct interpretation of UH coefficients, it is necessary to take into account that, as η_1 is set to zero, then

[9] Age covariates measure the difference between the current age (time-varying age) with respect to the legal working age in the Spanish labor market, 16 years-old.

[10] Nationality effect is captured using a dummy variable, that takes value one if the unemployed worker is not Spanish, to identify whether the unemployed worker is not Spanish.

[11] It refers to `Discrete time PH model without frailty` firstly estimated by `hshaz` and `hshaz2` commands, shown in pages 14 and 18 at Appendix.

[12] These coefficients are available at final Appendix, in pages 14 and 18.

η_2 estimates the differential unobserved effect (of being Type II unemployed workers) on the probability of exiting out of unemployment (state), with respect to the (estimated coefficient of the) regression constant term, -5.159466. Therefore, the estimated unobserved effect of being Type I and Type II unemployed workers are given by -5.159466 and -6.0860925 (=-5.159466-0.9266265), respectively.

```
. display "Started at $S_TIME"
Started at 19:19:53

. hshaz2 `varsaleU´ , id(codind) seq(j) d(exit) nmp(2) difficult
```

(output omitted)

```
Discrete time PH model, with discrete mixture     Number of obs   =     568042
                                                  LR chi2()       =          .
Log likelihood = -122078.61                       Prob > chi2     =          .
```

exit	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
hazard						
lnjunemp	3.614478	.0420878	85.88	0.000	3.531987	3.696968
lnjunemp2	-.9491475	.0099174	-95.71	0.000	-.9685853	-.9297097
month12	.1622685	.0306302	5.30	0.000	.1022344	.2223027
month18	.1378142	.051428	2.68	0.007	.0370172	.2386111
month24	.7349469	.0620328	11.85	0.000	.6133649	.8565289
ub	-1.28535	.0850207	-15.12	0.000	-1.451988	-1.118713
ubxlnjunemp	.5991264	.04521	13.25	0.000	.5105164	.6877363
female	-.2608072	.0143061	-18.23	0.000	-.2888466	-.2327679
age16tv	.1813952	.0079179	22.91	0.000	.1658763	.196914
age16tv2	-.0133666	.0007284	-18.35	0.000	-.0147943	-.011939
educcompul1	.090299	.0193612	4.66	0.000	.0523517	.1282463
educcompul2	.2061935	.0143468	14.37	0.000	.1780742	.2343127
inmigra	.1978054	.0221707	8.92	0.000	.1543517	.2412591
unrate	-.0783014	.0017806	-43.98	0.000	-.0817913	-.0748116
andal	.3975539	.0242093	16.42	0.000	.3501046	.4450033
aragon	-.3055231	.0429434	-7.11	0.000	-.3896905	-.2213556
astur	-.1925411	.0484716	-3.97	0.000	-.2875437	-.0975386
balear	.104185	.0369932	2.82	0.005	.0316797	.1766903
canar	.1056043	.0322674	3.27	0.001	.0423613	.1688473
cantab	-.1365828	.0591541	-2.31	0.021	-.2525228	-.0206428
castman	-.0313442	.0308195	-1.02	0.309	-.0917493	.029061
castleon	-.0867659	.0323628	-2.68	0.007	-.1501957	-.023336
valenc	.0931641	.0235731	3.95	0.000	.0469617	.1393665
extrem	.2210893	.0479023	4.62	0.000	.1272026	.3149761
galic	-.1072949	.0294785	-3.64	0.000	-.1650717	-.0495181
murcia	.0607897	.0389267	1.56	0.118	-.0155053	.1370846
navarr	-.3874108	.0714269	-5.42	0.000	-.5274051	-.2474166
vasco	-.2923917	.0420199	-6.96	0.000	-.3747493	-.2100341
rioja	-.1261293	.0740175	-1.70	0.088	-.2712009	.0189422
_cons	-5.159466	.0520456	-99.13	0.000	-5.261473	-5.057458
m2						
_cons	3.532398	.0322314	109.59	0.000	3.469226	3.59557
logitp2						
_cons	-.9266265	.0159517	-58.09	0.000	-.9578912	-.8953618
Prob. Type 1	.7163904	.003241	221.04	0.000	.7099954	.7226994
Prob. Type 2	.2836096	.003241	87.51	0.000	.2773006	.2900046

```
Note: m1 = 0

. display "Finished at $S_TIME"
Finished at 19:20:35
```

```
. display "Started at $S_TIME"
Started at 19:20:35

. hshaz `varsaleU´ , id(codind) seq(j) d(exit) nmp(2) difficult
  (output omitted)
Discrete time PH model, with discrete mixture     Number of obs   =     568042
                                                  LR chi2()       =          .
Log likelihood = -122078.61                       Prob > chi2     =          .

------------------------------------------------------------------------------
        exit |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hazard       |
    lnjunemp |   3.614478   .0420888    85.88   0.000     3.531985     3.69697
   lnjunemp2 |  -.9491475   .0099176   -95.70   0.000    -.9685857   -.9297093
     month12 |   .1622676   .0306303     5.30   0.000     .1022334    .2223018
     month18 |   .1378108    .051428     2.68   0.007     .0370137    .2386079
     month24 |   .7349434   .0620328    11.85   0.000     .6133612    .8565255
          ub |  -1.285337   .0850238   -15.12   0.000    -1.451981   -1.118693
 ubxlnjunemp |   .5991179   .0452117    13.25   0.000     .5105046    .6877312
      female |  -.2608074   .0143061   -18.23   0.000    -.2888467    -.232768
     age16tv |   .1813956   .0079181    22.91   0.000     .1658765    .1969147
    age16tv2 |  -.0133667   .0007284   -18.35   0.000    -.0147943    -.011939
 educcompul1 |   .0902986   .0193612     4.66   0.000     .0523513    .1282459
 educcompul2 |   .2061928   .0143468    14.37   0.000     .1780735    .2343121
     inmigra |   .1978042   .0221707     8.92   0.000     .1543505     .241258
      unrate |  -.0783014   .0017806   -43.98   0.000    -.0817912   -.0748116
       andal |   .3975506   .0242093    16.42   0.000     .3501011        .445
      aragon |  -.3055266   .0429434    -7.11   0.000    -.3896941   -.2213591
       astur |  -.1925455   .0484716    -3.97   0.000    -.2875481   -.0975429
      balear |   .1041814   .0369932     2.82   0.005     .0316761    .1766866
       canar |   .1055982   .0322675     3.27   0.001      .042355    .1688413
      cantab |  -.1365885   .0591542    -2.31   0.021    -.2525286   -.0206484
     castman |  -.0313481   .0308196    -1.02   0.309    -.0917533    .0290572
    castleon |  -.0867692   .0323628    -2.68   0.007     -.150199   -.0233393
      valenc |   .0931611   .0235731     3.95   0.000     .0469587    .1393635
      extrem |   .2210849   .0479023     4.62   0.000     .1271981    .3149717
       galic |  -.1072982   .0294785    -3.64   0.000     -.165075   -.0495214
      murcia |   .0607846   .0389268     1.56   0.118    -.0155104    .1370797
      navarr |  -.3874214   .0714272    -5.42   0.000    -.5274161   -.2474268
       vasco |  -.2923951     .04202    -6.96   0.000    -.3747527   -.2100375
       rioja |  -.1261426   .0740177    -1.70   0.088    -.2712147    .0189295
       _cons |  -5.159464   .0520466   -99.13   0.000    -5.261473   -5.057454
-------------+----------------------------------------------------------------
m2           |
       _cons |   3.532398   .0322319   109.59   0.000     3.469225    3.595571
-------------+----------------------------------------------------------------
logitp2      |
       _cons |  -.9266268   .0159517   -58.09   0.000    -.9578916   -.8953621
-------------+----------------------------------------------------------------
Prob. Type 1 |   .7163904    .003241   221.04   0.000     .7099955    .7226995
Prob. Type 2 |   .2836096    .003241    87.51   0.000     .2773005    .2900045
------------------------------------------------------------------------------
Note: m1 = 0

. display "Finished at $S_TIME"
Finished at 19:42:55
```

Table 1: Estimation of unispell duration models (Sample size: 568,042 observations)

	Time (hh:mm:ss)		
	`hshaz`		
	Start time	Finish time	Duration
Two mass-points	19:20:35	19:42:55	0:22:20
Three mass-points	19:44:08	20:27:16	0:43:08
	`hshaz2`		
	Start time	Finish time	Duration
Two mass-points	19:19:53	19:20:35	0:00:42
Three mass-points	19:42:55	19:44:08	0:01:13

Table 1 reports time spent by running both `hshaz` and `hshaz2` commands to achieve the convergence of the fitted duration models mentioned above, with two and three mass-points, respectively.[13] Results of Table 1 highlight two relevant differences between `hshaz2` and `hshaz` commands: First, `hshaz2` provides a significant reduction in time required to achieve the convengence: to fit a two mass-points model, `hshaz` spends thirty seven minutes and fifty five seconds minutes, while `hshaz2` spends only fifty one seconds, which means thirty seven minutes less. And, second, unlike `hshaz2`, time required by `hshaz` to achieve the convergence strongly depends on the number of points of support especified by command's user: `hshaz` needs one hour and fourteen minutes to fit a three mass-points model, while `hshaz2` spends only one minute and thirty one seconds.

In order to show the convergence process followed by running both `hshaz` and `hshaz2` commands, Figures 1 and 2 plot the log-likelihood values taken at each iteration by `hshaz` and `hshaz2` commands during the convergence process of the estimation of two mass-points and three mass-points unispell models, respectively. As can be seen in Figures 1 and 2, the values taken by the log-likelihood functions of `hshaz` and `hshaz2` at first iterations slightly differ, but when they aproximate to the maximum, the two log-likelihood functions converge to the same value: -122,078.61 for two mass-points models, and -121,861.5 for three mass-points models.

[13] For this specific sample of youth unemployed workers, the identification of unobserved heterogeneity is not possible when more than three mass-points are specified, which leads to not achieve convergence, neither running `hshaz`, nor `hshaz2` commands. This is the reason why only results from two and three mass-points duration models are shown in Table 1.

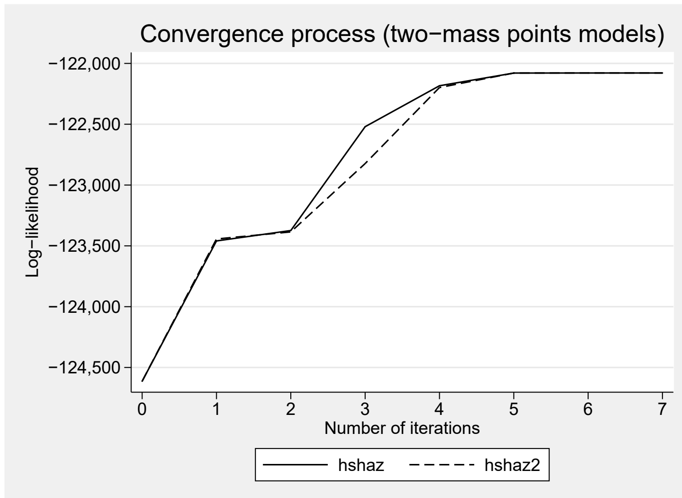

Figure 1: Iteration process of unispell duration models estimation (two mass-points)

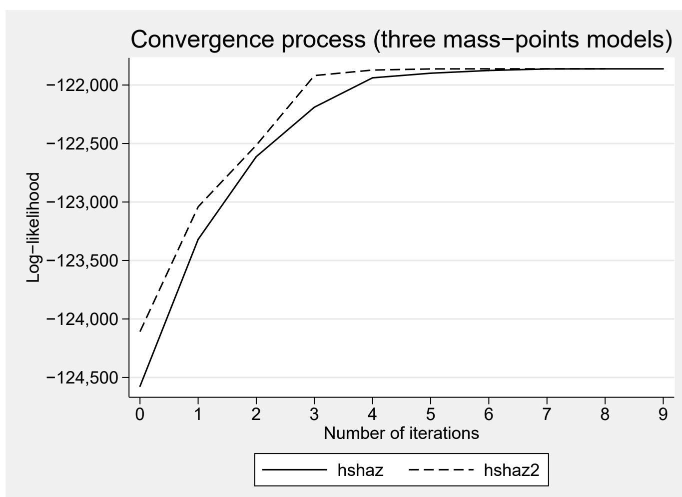

Figure 2: Iteration process of unispell duration models estimation (three mass-points)

Table 2: Estimation of multispell duration models using `hshaz2` (Sample size: 1,547,507 observations)

	Time (hh:mm:ss)		
	Start time	Finish time	Duration
Two mass-points	20:27:16	20:28:33	0:01:17
Three mass-points	20:28:33	20:30:50	0:02:17

5.2 Fitting multispell duration models using hshaz2

As was previously mentioned, `hshaz2` allows for the estimation of multispell duration models, by which individuals may be at risk of exiting more than once. And this feature implies, in our example, that each individual may experience several unemployment episodes. Time saving advantages given by `hshaz2` takes special interest in multispell duration models because of the increasing of the number of observations of the estimation sample, and therefore, the increase in required estimation time that it implies. To highlight the importance of this, a multispell duration model is fitted using another version of the sample, in which multiple unemployment spells are observed for each individual. This sample has 1,547,507 observations. The number of indivuals is 44,077, and the total number of unemployment spells is 146,851, that means an average number of spells per individual of roughly 3.33. The econometric especification includes the same covariates specified by the rest of the models estimated in the previous Section.

Table 2 reports time spent by `hshaz2` to achieve convergence for the two estimated models, with two and three points of support, respectively: the first one includes two points of support, and the second one specifies three points of support. Estimation output at final Appendix, in 7, reports detailed estimation output.

6 Reparameterization of mass-points probabilities

Functional form followed by `hshaz` command to compute mass-points probability parameters π_j is a *Logit*, regardless of the number of points of support specified by the command's user, whereas `hshaz2` computes the mass-points probabilities using a *Multinomial Logit.* For the estimation of two mass-points models, as only one parameter p_2 must be estimated, both `hshaz` and `hshaz2` compute the values of mass probability parameters using the same functional form, with $\pi_2 = \frac{e^{p2}}{1+e^{p2}}$ and $\pi_1 = 1 - \pi_2$, providing the same coefficient estimates of $\hat{\pi_2}$ and $\hat{\pi_1}$, as well as for their standard errors. However, when more than two points of support are especified, the functional forms used by `hshaz` and `hshaz2` to compute the values of mass probability parameters are different. For example, for three mass-points models, `hshaz` computes $\pi_2 = \frac{e^{p2}}{1+e^{p2}}$, $\pi_3 = \frac{e^{p3}}{1+e^{p3}}$

and $\pi_1 = 1-\pi_2-\pi_3$; whereas `hshaz2` computes $\pi_2 = \frac{e^{p2}}{1+e^{p2}+e^{p3}}$, $\pi_3 = \frac{e^{p3}}{1+e^{p2}+e^{p3}}$ and $\pi_1 = 1 - \pi_2 - \pi_3$.

This explians the differences found in mass-probability parameters estimates, given by $\hat{p}_2$, $\hat{p}_3$, $\hat{\pi}_2$ and $\hat{\pi}_3$, depending on whether have been estimated using `hshaz` or `hshaz2`. Thus, as can be seen in the estimation output of Appendix (in pages 18 to 21), estimated values of p_2 and p_3 provided by `hshaz` (`hshaz2`) are -0.915 (-0.629) and -1.533 (-1.105), respectively. And standard errors of p_2 and p_3 provided by `hshaz` (`hshaz2`) are 0.016 (0.025) and 0.081 (0.087), respectively. Both `hshaz` and `hshaz2` provide p_2 and p_3 coefficients statistically significant at 1% level.[14] However, estimated values of mass probability parameters, given by $\hat{p}_2$, $\hat{p}_3$ and $\hat{p}_1$, do not show significant differences.

7 Conclusion

`hshaz2` command provides the programming of gradient vector and Hessian matrix of the log-likelihood function of `hshaz` command, written by professor Stephen Jenkins. The programming of the algebraic expressions of both the first and second derivatives allows to use `d2 ml` method evaluator to achieve faster estimations of discrete time proportional duration models with unobserved heterogeneity. Furthermore, in contrast to `hshaz`, `hshaz2` allows for the estimation of multispell duration models, by which individuals may be observed at risk of exiting more than onece. The gains achieved in saving time provided `hshaz2` are stricky: Estimation results show that, on a sample size of 568,042 observations, `hshaz2` (`hshaz`) spends 0.42 (22.2) and 1.13 (43.08) minutes to achieve the convergence of a duration model with two and three points of support, respectively. Finally, The gains in estimation time involved by `hshaz2` command takes special relevance as the sample size increases: Using a sample with 1,547,507 observations, the estimation of a multispell duration model with two (three) points of support requires 1.17 (2.17) minutes.

[14] To estimate the standard errors of mass probability parameters, `hshaz2` provides to `diparm` command the algebraric expressions of first order derivatives of each $\pi_j = \frac{e^{p_j}}{1+\sum_{l=2}^{L} e^{p_l}}$, for each $j = 1, 2, \ldots, P$, with respect to each p_l, with $l = 2, 3, \ldots, P$.

References

[Gould et al., 2010] Gould, W., Pitblado, J. and Poi, B., *Maximum Likelihood Estimation with Stata*, Fourth Edition, Stata Press, 2010.

[Heckman and Singer, 1984] Heckman, J. J. and Singer, B., *A Method for Minimizing the Impact of the Distributional Assumptions in Econometric Models for Duration Data*, Econometrica, Vol. 52, pp. 271-320.

[Arranz and García-Serrano, 2011] Arranz, J.M. and García-Serrano, C. (2011) *Are the MCVL tax data useful? Ideas for mining*, Hacienda Pública Española, Vol. 199(4), 151-186.

[Lapuerta, 2010] Lapuerta, I. (2010) *Claves para el trabajo con la Muestra Continua de Vidas Laborales*, DemoSoc working paper (2010-37), Universitat Pompeu Fabra

[García-Pérez, 2008] García-Pérez, J.I., 2008 *La Muestra Continua de Vidas Laborales: Una guía de uso para el anlisis de transiciones*, Revista de Economía Aplicada, N. E-1, Vol. XVI, pp. 5-28.

[Jenkins, 2005] Jenkins, S., 2005 *Survival Analysis*, manuscript.

Appendix

Estimation output of fitting a two mass-points model using hshaz2 and hshaz

```
. display "Started at $S_TIME"
Started at 19:19:53

. hshaz2 `varsaleU´ , id(codind) seq(j) d(exit) nmp(2) difficult
Discrete time PH model without frailty

Generalized linear models                          No. of obs      =    568,042
Optimization     : ML                              Residual df     =    568,012
                                                   Scale parameter =          1
Deviance         =  247167.5906                    (1/df) Deviance =    .435145
Pearson          =  482484.0598                    (1/df) Pearson  =   .8494258

Variance function: V(u) = u*(1-u)                  [Bernoulli]
Link function    : g(u) = ln(-ln(1-u))             [Complementary log-log]

                                                   AIC             =   .4352277
Log likelihood   = -123583.7953                    BIC             =   -7278963

------------------------------------------------------------------------------
             |                 OIM
        exit |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnjunemp |   1.430829   .0203562    70.29   0.000     1.390932    1.470727
   lnjunemp2 |  -.5548032   .0060624   -91.52   0.000    -.5666853   -.5429211
     month12 |   .3093138   .0304787    10.15   0.000     .2495767    .3690509
     month18 |   .2920002   .0513613     5.69   0.000     .1913338    .3926665
     month24 |   .8134025   .0617714    13.17   0.000     .6923328    .9344722
          ub |  -.8271962   .0569521   -14.52   0.000    -.9388203   -.7155722
 ubxlnjunemp |   .4035376   .0374274    10.78   0.000     .3301811     .476894
      female |  -.2504213   .0113921   -21.98   0.000    -.2727495   -.2280931
     age16tv |   .1875119   .0060529    30.98   0.000     .1756485    .1993753
    age16tv2 |  -.0132609   .0005665   -23.41   0.000    -.0143712   -.0121507
 educcompul1 |   .0656452    .015286     4.29   0.000     .0356851    .0956053
 educcompul2 |   .1548147   .0113969    13.58   0.000     .1324771    .1771523
     inmigra |   .1929133   .0166177    11.61   0.000     .1603433    .2254834
      unrate |  -.0660709   .0014198   -46.53   0.000    -.0688537    -.063288
       andal |   .3343813   .0195089    17.14   0.000     .2961446    .3726181
      aragon |  -.2316286   .0333967    -6.94   0.000    -.2970849   -.1661722
       astur |  -.2065767   .0392823    -5.26   0.000    -.2835685   -.1295849
      balear |   .0027347   .0295148     0.09   0.926    -.0551133    .0605826
       canar |    .076421   .0259586     2.94   0.003     .0255431    .1272989
      cantab |   -.137379   .0488271    -2.81   0.005    -.2330784   -.0416796
     castman |  -.0116448   .0243639    -0.48   0.633    -.0593972    .0361075
    castleon |  -.0837173   .0262247    -3.19   0.001    -.1351167    -.032318
      valenc |   .1043138   .0185418     5.63   0.000     .0679726    .1406549
      extrem |   .1819348   .0379113     4.80   0.000       .10763    .2562395
       galic |  -.1072028   .0233738    -4.59   0.000    -.1530147   -.0613909
      murcia |   .0651937   .0305624     2.13   0.033     .0052925    .1250949
      navarr |  -.3191073   .0566289    -5.64   0.000    -.4300979   -.2081168
       vasco |  -.2487127   .0344305    -7.22   0.000    -.3161953   -.1812301
       rioja |   -.115534   .0606043    -1.91   0.057    -.2343161    .0032482
       _cons |  -2.454826    .024875   -98.69   0.000     -2.50358   -2.406072
------------------------------------------------------------------------------
```

(output omitted)

```
Discrete time PH model, with discrete mixture     Number of obs   =     568042
                                                  LR chi2()       =          .
Log likelihood = -122078.61                       Prob > chi2     =          .
```

exit	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
hazard						
lnjunemp	3.614478	.0420878	85.88	0.000	3.531987	3.696968
lnjunemp2	-.9491475	.0099174	-95.71	0.000	-.9685853	-.9297097
month12	.1622685	.0306302	5.30	0.000	.1022344	.2223027
month18	.1378142	.051428	2.68	0.007	.0370172	.2386111
month24	.7349469	.0620328	11.85	0.000	.6133649	.8565289
ub	-1.28535	.0850207	-15.12	0.000	-1.451988	-1.118713
ubxlnjunemp	.5991264	.04521	13.25	0.000	.5105164	.6877363
female	-.2608072	.0143061	-18.23	0.000	-.2888466	-.2327679
age16tv	.1813952	.0079179	22.91	0.000	.1658763	.196914
age16tv2	-.0133666	.0007284	-18.35	0.000	-.0147943	-.011939
educcompul1	.090299	.0193612	4.66	0.000	.0523517	.1282463
educcompul2	.2061935	.0143468	14.37	0.000	.1780742	.2343127
inmigra	.1978054	.0221707	8.92	0.000	.1543517	.2412591
unrate	-.0783014	.0017806	-43.98	0.000	-.0817913	-.0748116
andal	.3975539	.0242093	16.42	0.000	.3501046	.4450033
aragon	-.3055231	.0429434	-7.11	0.000	-.3896905	-.2213556
astur	-.1925411	.0484716	-3.97	0.000	-.2875437	-.0975386
balear	.104185	.0369932	2.82	0.005	.0316797	.1766903
canar	.1056043	.0322674	3.27	0.001	.0423613	.1688473
cantab	-.1365828	.0591541	-2.31	0.021	-.2525228	-.0206428
castman	-.0313442	.0308195	-1.02	0.309	-.0917493	.029061
castleon	-.0867659	.0323628	-2.68	0.007	-.1501957	-.023336
valenc	.0931641	.0235731	3.95	0.000	.0469617	.1393665
extrem	.2210893	.0479023	4.62	0.000	.1272026	.3149761
galic	-.1072949	.0294785	-3.64	0.000	-.1650717	-.0495181
murcia	.0607897	.0389267	1.56	0.118	-.0155053	.1370846
navarr	-.3874108	.0714269	-5.42	0.000	-.5274051	-.2474166
vasco	-.2923917	.0420199	-6.96	0.000	-.3747493	-.2100341
rioja	-.1261293	.0740175	-1.70	0.088	-.2712009	.0189422
_cons	-5.159466	.0520456	-99.13	0.000	-5.261473	-5.057458
m2						
_cons	3.532398	.0322314	109.59	0.000	3.469226	3.59557
logitp2						
_cons	-.9266265	.0159517	-58.09	0.000	-.9578912	-.8953618
Prob. Type 1	.7163904	.003241	221.04	0.000	.7099954	.7226994
Prob. Type 2	.2836096	.003241	87.51	0.000	.2773006	.2900046

```
Note: m1 = 0

. display "Finished at $S_TIME"
Finished at 19:20:35
```

```
. display "Started at $S_TIME"
Started at 19:20:35

. hshaz `varsaleU´ , id(codind) seq(j) d(exit) nmp(2) difficult
Discrete time PH model without frailty

Generalized linear models                          No. of obs      =    568,042
Optimization     : ML                              Residual df     =    568,012
                                                   Scale parameter =          1
Deviance         =  247167.5906                    (1/df) Deviance =    .435145
Pearson          =  482484.0598                    (1/df) Pearson  =   .8494258

Variance function: V(u) = u*(1-u)                  [Bernoulli]
Link function    : g(u) = ln(-ln(1-u))             [Complementary log-log]

                                                   AIC             =   .4352277
Log likelihood   = -123583.7953                    BIC             =   -7278963

------------------------------------------------------------------------------
             |                 OIM
        exit |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnjunemp |   1.430829   .0203562    70.29   0.000     1.390932    1.470727
   lnjunemp2 |  -.5548032   .0060624   -91.52   0.000    -.5666853   -.5429211
     month12 |   .3093138   .0304787    10.15   0.000     .2495767    .3690509
     month18 |   .2920002   .0513613     5.69   0.000     .1913338    .3926665
     month24 |   .8134025   .0617714    13.17   0.000     .6923328    .9344722
          ub |  -.8271962   .0569521   -14.52   0.000    -.9388203   -.7155722
 ubxlnjunemp |   .4035376   .0374274    10.78   0.000     .3301811     .476894
      female |  -.2504213   .0113921   -21.98   0.000    -.2727495   -.2280931
     age16tv |   .1875119   .0060529    30.98   0.000     .1756485    .1993753
    age16tv2 |  -.0132609   .0005665   -23.41   0.000    -.0143712   -.0121507
 educcompul1 |   .0656452    .015286     4.29   0.000     .0356851    .0956053
 educcompul2 |   .1548147   .0113969    13.58   0.000     .1324771    .1771523
     inmigra |   .1929133   .0166177    11.61   0.000     .1603433    .2254834
      unrate |  -.0660709   .0014198   -46.53   0.000    -.0688537    -.063288
       andal |   .3343813   .0195089    17.14   0.000     .2961446    .3726181
      aragon |  -.2316286   .0333967    -6.94   0.000    -.2970849   -.1661722
       astur |  -.2065767   .0392823    -5.26   0.000    -.2835685   -.1295849
      balear |   .0027347   .0295148     0.09   0.926    -.0551133    .0605826
       canar |    .076421   .0259586     2.94   0.003     .0255431    .1272989
      cantab |   -.137379   .0488271    -2.81   0.005    -.2330784   -.0416796
     castman |  -.0116448   .0243639    -0.48   0.633    -.0593972    .0361075
    castleon |  -.0837173   .0262247    -3.19   0.001    -.1351167    -.032318
      valenc |   .1043138   .0185418     5.63   0.000     .0679726    .1406549
      extrem |   .1819348   .0379113     4.80   0.000       .10763    .2562395
       galic |  -.1072028   .0233738    -4.59   0.000    -.1530147   -.0613909
      murcia |   .0651937   .0305624     2.13   0.033     .0052925    .1250949
      navarr |  -.3191073   .0566289    -5.64   0.000    -.4300979   -.2081168
       vasco |  -.2487127   .0344305    -7.22   0.000    -.3161953   -.1812301
       rioja |   -.115534   .0606043    -1.91   0.057    -.2343161    .0032482
       _cons |  -2.454826    .024875   -98.69   0.000     -2.50358   -2.406072
------------------------------------------------------------------------------
```

(output omitted)

```
Discrete time PH model, with discrete mixture     Number of obs   =     568042
                                                  LR chi2()       =          .
Log likelihood = -122078.61                       Prob > chi2     =          .
```

exit	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
hazard						
lnjunemp	3.614478	.0420888	85.88	0.000	3.531985	3.69697
lnjunemp2	-.9491475	.0099176	-95.70	0.000	-.9685857	-.9297093
month12	.1622676	.0306303	5.30	0.000	.1022334	.2223018
month18	.1378108	.051428	2.68	0.007	.0370137	.2386079
month24	.7349434	.0620328	11.85	0.000	.6133612	.8565255
ub	-1.285337	.0850238	-15.12	0.000	-1.451981	-1.118693
ubxlnjunemp	.5991179	.0452117	13.25	0.000	.5105046	.6877312
female	-.2608074	.0143061	-18.23	0.000	-.2888467	-.232768
age16tv	.1813956	.0079181	22.91	0.000	.1658765	.1969147
age16tv2	-.0133667	.0007284	-18.35	0.000	-.0147943	-.011939
educcompul1	.0902986	.0193612	4.66	0.000	.0523513	.1282459
educcompul2	.2061928	.0143468	14.37	0.000	.1780735	.2343121
inmigra	.1978042	.0221707	8.92	0.000	.1543505	.241258
unrate	-.0783014	.0017806	-43.98	0.000	-.0817912	-.0748116
andal	.3975506	.0242093	16.42	0.000	.3501011	.445
aragon	-.3055266	.0429434	-7.11	0.000	-.3896941	-.2213591
astur	-.1925455	.0484716	-3.97	0.000	-.2875481	-.0975429
balear	.1041814	.0369932	2.82	0.005	.0316761	.1766866
canar	.1055982	.0322675	3.27	0.001	.042355	.1688413
cantab	-.1365885	.0591542	-2.31	0.021	-.2525286	-.0206484
castman	-.0313481	.0308196	-1.02	0.309	-.0917533	.0290572
castleon	-.0867692	.0323628	-2.68	0.007	-.150199	-.0233393
valenc	.0931611	.0235731	3.95	0.000	.0469587	.1393635
extrem	.2210849	.0479023	4.62	0.000	.1271981	.3149717
galic	-.1072982	.0294785	-3.64	0.000	-.165075	-.0495214
murcia	.0607846	.0389268	1.56	0.118	-.0155104	.1370797
navarr	-.3874214	.0714272	-5.42	0.000	-.5274161	-.2474268
vasco	-.2923951	.04202	-6.96	0.000	-.3747527	-.2100375
rioja	-.1261426	.0740177	-1.70	0.088	-.2712147	.0189295
_cons	-5.159464	.0520466	-99.13	0.000	-5.261473	-5.057454
m2						
_cons	3.532398	.0322319	109.59	0.000	3.469225	3.595571
logitp2						
_cons	-.9266268	.0159517	-58.09	0.000	-.9578916	-.8953621
Prob. Type 1	.7163904	.003241	221.04	0.000	.7099955	.7226995
Prob. Type 2	.2836096	.003241	87.51	0.000	.2773005	.2900045

```
Note: m1 = 0

. display "Finished at $S_TIME"
Finished at 19:42:55
```

Estimation output of fitting a three mass-points model using hshaz2 and hshaz

```
. display "Started at $S_TIME"
Started at 19:42:55

. hshaz2 `varsaleU´ , id(codind) seq(j) d(exit) nmp(3) difficult
Discrete time PH model without frailty

Generalized linear models                          No. of obs      =    568,042
Optimization     : ML                              Residual df     =    568,012
                                                   Scale parameter =          1
Deviance         =  247167.5906                    (1/df) Deviance =    .435145
Pearson          =  482484.0598                    (1/df) Pearson  =   .8494258

Variance function: V(u) = u*(1-u)                  [Bernoulli]
Link function    : g(u) = ln(-ln(1-u))             [Complementary log-log]

                                                   AIC             =   .4352277
Log likelihood   = -123583.7953                    BIC             =   -7278963

------------------------------------------------------------------------------
             |                 OIM
        exit |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnjunemp |   1.430829   .0203562    70.29   0.000     1.390932    1.470727
   lnjunemp2 |  -.5548032   .0060624   -91.52   0.000    -.5666853   -.5429211
     month12 |   .3093138   .0304787    10.15   0.000     .2495767    .3690509
     month18 |   .2920002   .0513613     5.69   0.000     .1913338    .3926665
     month24 |   .8134025   .0617714    13.17   0.000     .6923328    .9344722
          ub |  -.8271962   .0569521   -14.52   0.000    -.9388203   -.7155722
 ubxlnjunemp |   .4035376   .0374274    10.78   0.000     .3301811     .476894
      female |  -.2504213   .0113921   -21.98   0.000    -.2727495   -.2280931
     age16tv |   .1875119   .0060529    30.98   0.000     .1756485    .1993753
    age16tv2 |  -.0132609   .0005665   -23.41   0.000    -.0143712   -.0121507
 educcompul1 |   .0656452    .015286     4.29   0.000     .0356851    .0956053
 educcompul2 |   .1548147   .0113969    13.58   0.000     .1324771    .1771523
     inmigra |   .1929133   .0166177    11.61   0.000     .1603433    .2254834
      unrate |  -.0660709   .0014198   -46.53   0.000    -.0688537    -.063288
       andal |   .3343813   .0195089    17.14   0.000     .2961446    .3726181
      aragon |  -.2316286   .0333967    -6.94   0.000    -.2970849   -.1661722
       astur |  -.2065767   .0392823    -5.26   0.000    -.2835685   -.1295849
      balear |   .0027347   .0295148     0.09   0.926    -.0551133    .0605826
       canar |    .076421   .0259586     2.94   0.003     .0255431    .1272989
      cantab |   -.137379   .0488271    -2.81   0.005    -.2330784   -.0416796
     castman |  -.0116448   .0243639    -0.48   0.633    -.0593972    .0361075
    castleon |  -.0837173   .0262247    -3.19   0.001    -.1351167    -.032318
      valenc |   .1043138   .0185418     5.63   0.000     .0679726    .1406549
      extrem |   .1819348   .0379113     4.80   0.000       .10763    .2562395
       galic |  -.1072028   .0233738    -4.59   0.000    -.1530147   -.0613909
      murcia |   .0651937   .0305624     2.13   0.033     .0052925    .1250949
      navarr |  -.3191073   .0566289    -5.64   0.000    -.4300979   -.2081168
       vasco |  -.2487127   .0344305    -7.22   0.000    -.3161953   -.1812301
       rioja |   -.115534   .0606043    -1.91   0.057    -.2343161    .0032482
       _cons |  -2.454826    .024875   -98.69   0.000     -2.50358   -2.406072
------------------------------------------------------------------------------
```

(output omitted)

```
Discrete time PH model, with discrete mixture     Number of obs   =     568042
                                                  LR chi2()       =          .
Log likelihood =  -121861.5                       Prob > chi2     =          .
```

exit	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
hazard						
lnjunemp	3.498714	.0440098	79.50	0.000	3.412457	3.584972
lnjunemp2	-.8344097	.0114636	-72.79	0.000	-.856878	-.8119415
month12	.1728664	.0310214	5.57	0.000	.1120656	.2336672
month18	.171937	.0517658	3.32	0.001	.0704779	.2733961
month24	.8131917	.0624735	13.02	0.000	.6907459	.9356375
ub	-1.369189	.08402	-16.30	0.000	-1.533865	-1.204512
ubxlnjunemp	.544226	.0486218	11.19	0.000	.4489291	.6395229
female	-.3663279	.0192127	-19.07	0.000	-.403984	-.3286717
age16tv	.2851709	.0104771	27.22	0.000	.2646363	.3057056
age16tv2	-.0195095	.0009353	-20.86	0.000	-.0213427	-.0176763
educcompul1	.1166009	.0260306	4.48	0.000	.065582	.1676199
educcompul2	.2443288	.0190789	12.81	0.000	.2069349	.2817227
inmigra	.4532836	.0349271	12.98	0.000	.3848278	.5217394
unrate	-.1002291	.0023965	-41.82	0.000	-.1049261	-.0955322
andal	.4951261	.0331679	14.93	0.000	.4301183	.5601339
aragon	-.3788273	.0558053	-6.79	0.000	-.4882037	-.269451
astur	-.2838207	.0657935	-4.31	0.000	-.4127736	-.1548679
balear	.1441218	.0492659	2.93	0.003	.0475624	.2406811
canar	.0761471	.0427015	1.78	0.075	-.0075462	.1598404
cantab	-.1574583	.0770308	-2.04	0.041	-.3084358	-.0064807
castman	.0211891	.042082	0.50	0.615	-.0612901	.1036684
castleon	-.1217481	.0420095	-2.90	0.004	-.2040852	-.039411
valenc	.1208921	.0317378	3.81	0.000	.0586872	.1830971
extrem	.2780232	.0625915	4.44	0.000	.1553461	.4007003
galic	-.1282771	.0394518	-3.25	0.001	-.2056012	-.050953
murcia	.0963948	.0517655	1.86	0.063	-.0050638	.1978534
navarr	-.514514	.0934307	-5.51	0.000	-.6976348	-.3313933
vasco	-.4258417	.0551267	-7.72	0.000	-.533888	-.3177953
rioja	-.1639572	.1025683	-1.60	0.110	-.3649875	.037073
_cons	-4.985959	.0561667	-88.77	0.000	-5.096044	-4.875874
m2						
_cons	3.33153	.0341016	97.69	0.000	3.264692	3.398368
m3						
_cons	-1.853107	.0678011	-27.33	0.000	-1.985995	-1.72022
logitp2						
_cons	-.6293778	.0251716	-25.00	0.000	-.6787132	-.5800423
logitp3						
_cons	-1.105979	.0877127	-12.61	0.000	-1.277892	-.9340649
Prob. Type 1	.5365353	.011589	46.30	0.000	.5137611	.5591582
Prob. Type 2	.2859322	.0033266	85.95	0.000	.279457	.2924965
Prob. Type 3	.1775325	.0118536	14.98	0.000	.155478	.2019673

```
Note: m1 = 0

. display "Finished at $S_TIME"
Finished at 19:44:08
```

```
. display "Started at $S_TIME"
Started at 19:44:08

. hshaz `varsaleU´ , id(codind) seq(j) d(exit) nmp(3) difficult
Discrete time PH model without frailty

Generalized linear models                          No. of obs      =    568,042
Optimization     : ML                              Residual df     =    568,012
                                                   Scale parameter =          1
Deviance         =  247167.5906                    (1/df) Deviance =    .435145
Pearson          =  482484.0598                    (1/df) Pearson  =   .8494258

Variance function: V(u) = u*(1-u)                  [Bernoulli]
Link function    : g(u) = ln(-ln(1-u))             [Complementary log-log]

                                                   AIC             =   .4352277
Log likelihood   = -123583.7953                    BIC             =   -7278963

------------------------------------------------------------------------------
             |                 OIM
        exit |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnjunemp |   1.430829   .0203562    70.29   0.000     1.390932    1.470727
   lnjunemp2 |  -.5548032   .0060624   -91.52   0.000    -.5666853   -.5429211
     month12 |   .3093138   .0304787    10.15   0.000     .2495767    .3690509
     month18 |   .2920002   .0513613     5.69   0.000     .1913338    .3926665
     month24 |   .8134025   .0617714    13.17   0.000     .6923328    .9344722
          ub |  -.8271962   .0569521   -14.52   0.000    -.9388203   -.7155722
 ubxlnjunemp |   .4035376   .0374274    10.78   0.000     .3301811     .476894
      female |  -.2504213   .0113921   -21.98   0.000    -.2727495   -.2280931
     age16tv |   .1875119   .0060529    30.98   0.000     .1756485    .1993753
    age16tv2 |  -.0132609   .0005665   -23.41   0.000    -.0143712   -.0121507
 educcompul1 |   .0656452    .015286     4.29   0.000     .0356851    .0956053
 educcompul2 |   .1548147   .0113969    13.58   0.000     .1324771    .1771523
     inmigra |   .1929133   .0166177    11.61   0.000     .1603433    .2254834
      unrate |  -.0660709   .0014198   -46.53   0.000    -.0688537    -.063288
       andal |   .3343813   .0195089    17.14   0.000     .2961446    .3726181
      aragon |  -.2316286   .0333967    -6.94   0.000    -.2970849   -.1661722
       astur |  -.2065767   .0392823    -5.26   0.000    -.2835685   -.1295849
      balear |   .0027347   .0295148     0.09   0.926    -.0551133    .0605826
       canar |    .076421   .0259586     2.94   0.003     .0255431    .1272989
      cantab |   -.137379   .0488271    -2.81   0.005    -.2330784   -.0416796
     castman |  -.0116448   .0243639    -0.48   0.633    -.0593972    .0361075
    castleon |  -.0837173   .0262247    -3.19   0.001    -.1351167    -.032318
      valenc |   .1043138   .0185418     5.63   0.000     .0679726    .1406549
      extrem |   .1819348   .0379113     4.80   0.000       .10763    .2562395
       galic |  -.1072028   .0233738    -4.59   0.000    -.1530147   -.0613909
      murcia |   .0651937   .0305624     2.13   0.033     .0052925    .1250949
      navarr |  -.3191073   .0566289    -5.64   0.000    -.4300979   -.2081168
       vasco |  -.2487127   .0344305    -7.22   0.000    -.3161953   -.1812301
       rioja |   -.115534   .0606043    -1.91   0.057    -.2343161    .0032482
       _cons |  -2.454826    .024875   -98.69   0.000     -2.50358   -2.406072
------------------------------------------------------------------------------
```

(*output omitted*)

```
Discrete time PH model, with discrete mixture     Number of obs   =     568042
                                                  LR chi2()       =          .
Log likelihood =  -121861.5                       Prob > chi2     =          .

------------------------------------------------------------------------------
        exit |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hazard       |
    lnjunemp |   3.498702   .0440091    79.50   0.000     3.412446    3.584959
   lnjunemp2 |  -.8344044   .0114635   -72.79   0.000    -.8568725   -.8119363
     month12 |   .1728618   .0310213     5.57   0.000     .1120613    .2336623
     month18 |   .1719335   .0517656     3.32   0.001     .0704749    .2733922
     month24 |   .8131908   .0624731    13.02   0.000     .6907457    .9356359
          ub |  -1.369188   .0840199   -16.30   0.000    -1.533864   -1.204512
 ubxlnjunemp |   .5442215   .0486217    11.19   0.000     .4489248    .6395182
      female |   -.36633    .0192127   -19.07   0.000    -.4039861   -.3286739
     age16tv |   .2851726   .0104772    27.22   0.000     .2646376    .3057075
    age16tv2 |  -.0195095   .0009353   -20.86   0.000    -.0213428   -.0176763
 educcompul1 |   .1166015   .0260305     4.48   0.000     .0655826    .1676204
 educcompul2 |    .244328   .0190788    12.81   0.000     .2069343    .2817218
     inmigra |   .4532945    .034927    12.98   0.000     .3848389    .5217501
      unrate |  -.1002293   .0023965   -41.82   0.000    -.1049263   -.0955324
       andal |   .4951238   .0331676    14.93   0.000     .4301164    .5601312
      aragon |  -.3788299   .0558052    -6.79   0.000    -.4882061   -.2694537
       astur |   -.283826   .0657933    -4.31   0.000    -.4127785   -.1548735
      balear |   .1441224   .0492658     2.93   0.003     .0475631    .2406817
       canar |   .0761416   .0427012     1.78   0.075    -.0075511    .1598343
      cantab |  -.1574604   .0770308    -2.04   0.041    -.3084379   -.0064829
     castman |   .0211906    .042082     0.50   0.615    -.0612887    .1036699
    castleon |  -.1217498   .0420094    -2.90   0.004    -.2040866   -.0394129
      valenc |   .1208901   .0317377     3.81   0.000     .0586855    .1830948
      extrem |     .27802   .0625916     4.44   0.000     .1553428    .4006973
       galic |   -.128278   .0394517    -3.25   0.001     -.205602    -.050954
      murcia |   .0963935   .0517655     1.86   0.063     -.005065    .1978519
      navarr |   -.514519   .0934308    -5.51   0.000    -.6976399   -.3313981
       vasco |  -.4258472   .0551265    -7.72   0.000    -.5338931   -.3178014
       rioja |  -.1639663   .1025687    -1.60   0.110    -.3649972    .0370647
       _cons |  -4.985975   .0561658   -88.77   0.000    -5.096058   -4.875892
-------------+----------------------------------------------------------------
m2           |
       _cons |   3.331543   .0341009    97.70   0.000     3.264706    3.398379
-------------+----------------------------------------------------------------
m3           |
       _cons |  -1.853159   .0677947   -27.33   0.000    -1.986034   -1.720284
-------------+----------------------------------------------------------------
logitp2      |
       _cons |   -.915219   .0162927   -56.17   0.000    -.9471521   -.8832858
-------------+----------------------------------------------------------------
logitp3      |
       _cons |  -1.533228   .0811708   -18.89   0.000     -1.69232   -1.374136
-------------+----------------------------------------------------------------
Prob. Type 1 |    .536545   .0115873    46.30   0.000     .5137742    .5591644
Prob. Type 2 |   .2859331   .0033266    85.95   0.000     .2794579    .2924973
Prob. Type 3 |   .1775219   .0118516    14.98   0.000      .155471    .2019525
------------------------------------------------------------------------------
Note: m1 = 0

. display "Finished at $S_TIME"
Finished at 20:27:16
```

ÚLTIMOS DOCUMENTOS DE TRABAJO

2017-06: "Faster estimation of discrete time duration models with unobserved heterogeneity using hshaz2", **David Troncoso Ponce**.
2017-05: "Heterogeneous Household Finances and the Effect of Fiscal Policy, Javier Andrés,José E.Boscá, Javier Ferri y Cristina Fuentes-Albero.
2017-04: "Statistical Discrimination and the Efficiency of Quotas",J.Ignacio Conde-Ruiz,Juan-José Ganuza y Paola Profeta
2017-03: "Cargos por Azar", **Emilio Albi**.
2017-02: "Should pensions be redistributive? The impact of Spanish reforms on the system's sustainability and adequacy", **Concepció Patxot, Meritxell Solé and Guadalupe Souto**.
2017-01: "El Modelo de Perfilado Estadístico: una herramienta eficiente para caracterizar a los demandantes de empleo", **Yolanda F. Rebollo-Sanz**.
2016-10: "Family Job Search and Wealth: The Added Worker Effect Revisited", **J. Ignacio García-Pérez y Sílvio Rendon**.
2016-09: "Evolución del Gasto Público por Funciones durante la crisis (2007-2014): España vs UE", **José Ignacio Conde-Ruiz, Manuel Díaz , Carmen Marín y Juan F. Rubio-Ramírez**.
2016-08: "Thinking of Incentivizing Care? The Effect of Demand Subsidies on Informal Caregiving and Intergenerational Transfers", **Joan Costa-Font, Sergi Jiménez-Martín y Cristina Vilaplana-Prieto**.
2016-07: "The Pruned State-Space System for Non-Linear DSGE Models: Theory and Empirical Applications", **Martin M. Andreasen, Jesús Fernández-Villaverde y Juan F. Rubio-Ramírez**.
2016-06: "The effects of non-adherence on health care utilisation: panel data evidence on uncontrolled diabetes", **Joan Gil, Antonio Sicras-Mainar y Eugenio Zucchelli**.
2016-05: "Does Long-Term Care Subsidisation Reduce Unnecessary Hospitalisations?", **Joan Costa-Font, Sergi Jiménez-Martín y Cristina Vilaplana-Prieto**
2016-04: ""Cultural Persistence" of Health Capital: Evidence from European Migrants", **Joan Costa-Font y Azusa Sato**.
2016-03: "Like Mother, Like Father? Gender Assortative Transmission Of Child Overweight", **Joan Costa-Font y Mireia Jofre-Bonet**.
2016-02: "Health Capacity to Work at Older Ages: Evidence from Spain", **Pilar García-Gómez, Sergi Jimenez Martin y Judit Vall Castello**.
2016-01: "Monte Carlo evidence on the estimation of AR(1) panel data sample selection models", **Sergi Jiménez-Martín y José María Labeaga**.
2015-13: "On the Treatment of Foreigners and Foreign-Owned Firms in Cost–Benefit Analysis", **Per-Olov Johansson y Ginés de Rus**.
2015-12: "Evaluating Options for Shifting Tax Burden to Top Income Earners", **Jorge Onrubia, Fidel Picos y María del Carmen Rodado**.
2015-11: "Differences in Job De-Routinization in OECD countries: Evidence from PIAAC", **Sara De La Rica y Lucas Gortazar**.
2015-10: "Bad times, slimmer children?", **Cristina Belles-Obrero, Sergi Jimenez-Martín y Judit Vall-Castello**.
2015-09: "The Unintended Effects of Increasing the Legal Working Age on Family Behaviour", **Cristina Belles-Obrero, Sergi Jimenez-Martín y Judit Vall-Castello**.
2015-08: "Capital Humano y Productividad", **Ángel de la Fuente**.
2015-07: "The effect of changes in the statutory minimum working age on educational, labor and health outcomes", **Sergi Jiménez-Martín, Judit Vall y Elena del Rey**.
2015-06: "The Effects of Employment Uncertainty, Unemployment Insurance, and Wealth Shocks on the Retirement Behavior of Older Americans", **Hugo Benítez-Silva, J. Ignacio García-Pérez y Sergi Jiménez-Martín**.
2015-05: "Instruments, rules and household debt: The effects of fiscal policy", **J. Andrés, J.E. Boscá y J. Ferri**.
2015-04: "Can International Macroeconomic Models Explain Low-Frequency Movements of Real Exchange Rates?", **Pau Rabanal y Juan F. Rubio-Ramírez**.
2015-03: "Privatización, competencia y regulación aeroportuaria: Experiencia internacional", **Ofelia Betancor y María Paz Espinosa**.
2015-02: "La experiencia internacional en alta velocidad ferroviaria", **Daniel Albalate y Germà Bel**.
2015-01: "Household Debt and Fiscal Multipliers", **J. Andrés, J.E. Boscá y J. Ferri**.
2014-21: "Structural Estimation of a Model of School Choices: the Boston Mechanism vs. Its Alternatives", **Caterina Calsamiglia, Chao Fu y Maia Güell**.
2014-20: "Which club should I attend, Dad?: Targeted socialization and production", **Facundo Albornoz, Antonio Cabrales y Esther Hauk**.
2014-19: "The Informational Content of Surnames, the Evolution of Intergenerational Mobility and Assortative **Mating", Maia Güell, José V. Rodríguez Mora y Chris Telmer**.
2014-18: "Risk-sharing and contagion in networks", **Antonio Cabrales, Piero Gottardi y Fernando Vega-Redondo**.