Working papers series

WP ECON 17.05

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

David Troncoso-Ponce

Department of Economics, Universidad Pablo de Olavide

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

JEL Classification: C41, J64

Department of Economics

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 multispell 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, higher 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 Section 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 $\pi_j = \frac{e^{p_j}}{1+\sum_{l=2}^{P} e^{p_l}}$, $j = 2, 3, \ldots, P$. Each π_j parameter estimates the probability of observing each type of workers in 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 model parameters. The main contribution of `hshaz2` command is that it provides the algrebraic expressions of both first and second order 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. 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.

Coefficient estimates of model with UH show positive duration dependence ($Log(t) = 3.614$), but this positive effect decreases ($Log(t)^2 = -0.949$) as time spent in current unemployment state increases. Moreover, comparison between coefficient estimates with and without UH reveals the importance of controlling for the presence of UH. Thus, coefficient estimates of two mass-point duration model that does not control the presence of UH underestimates ($Log(t) = 1.43$ and $Log(t)^2 = -0.554$) the effect of unemployment duration dependence.

Regarding the estimation of the unobserved heterogeneity distribution, 71.6% of the sample are Type I unemployed workers, and the rest 28.4% 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 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 estimated coefficient of the regression constant term, -5.159. Therefore, the estimated unobserved effect of being Type I and Type II unemployed workers are given by -5.159 and -6.085 ($= -5.159 - 0.926$), respectively.

[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.

```
. 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.[11] 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 twenty two minutes and twenty seconds, while `hshaz2` spends only fourty two seconds, which means twenty eight 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 fourty three minutes to fit a three mass-points model, while `hshaz2` spends only one minute and thirteen 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.

[11] 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.

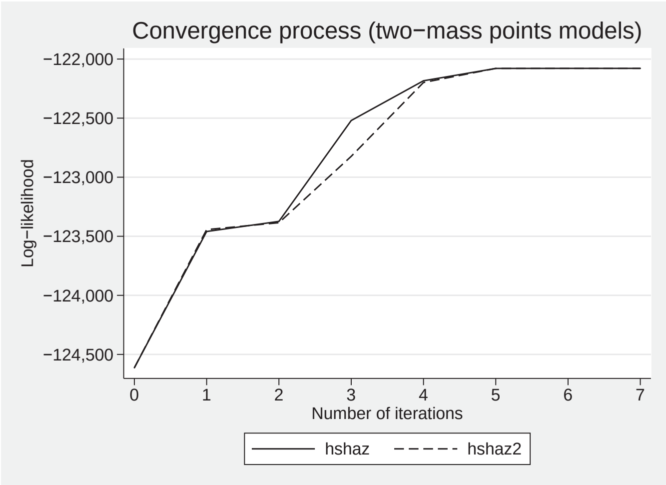

Figure 1: Iteration process of unispell duration models estimation (two mass-points)

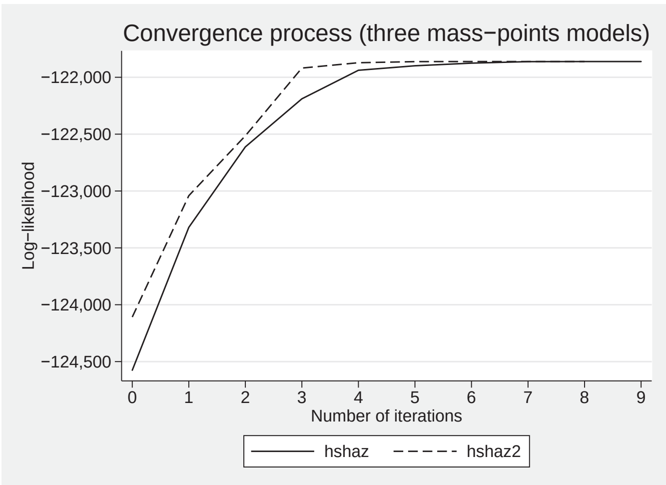

Figure 2: Iteration process of unispell duration models estimation (three mass-points)

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 increase 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: the first one includes two points of support, and the second one specifies three points of support. Estimation output at final Appendix 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, 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.[12] 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 first and second derivatives allows to use `d2 ml` method to achieve faster estimations of discrete time proportional hazard 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 once. The gains achieved in saving time provided by `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.

[12] 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 an unispell 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
```

exit	Coef.	OIM 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
```

exit	Coef.	OIM 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 an unispell 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
```

exit	Coef.	OIM 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
```

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

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

. hshaz3 `varsaleU´ , id(codind) spell(spell) seq(j) d(exit) nmp(2) difficult
Discrete time PH model without frailty

Generalized linear models                          No. of obs      =    1547507
Optimization     : ML                              Residual df     =    1547477
                                                   Scale parameter =          1
Deviance         =  787319.2776                    (1/df) Deviance =   .5087761
Pearson          =  1339880.771                    (1/df) Pearson  =   .8658486

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

                                                   AIC             =    .508805
Log likelihood   = -393659.6388                    BIC             =  -2.13e+07
```

exit	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
lnjunemp	1.529216	.0111456	137.20	0.000	1.507371	1.551061
lnjunemp2	-.6168131	.0035334	-174.57	0.000	-.6237384	-.6098879
month12	.2412328	.0188906	12.77	0.000	.2042079	.2782578
month18	.3530975	.0319817	11.04	0.000	.2904145	.4157804
month24	.8596715	.0405118	21.22	0.000	.7802699	.9390731
ub	-.9504732	.0217277	-43.74	0.000	-.9930587	-.9078876
ubxlnjunemp	.432762	.014572	29.70	0.000	.4042014	.4613226
female	-.1934465	.0061679	-31.36	0.000	-.2055354	-.1813577
age16tv	.1308036	.0025865	50.57	0.000	.1257342	.1358729
age16tv2	-.0071448	.0001769	-40.39	0.000	-.0074914	-.0067981
educcompul1	.0456268	.0081124	5.62	0.000	.0297268	.0615268
educcompul2	.1441138	.0061536	23.42	0.000	.1320529	.1561747
inmigra	.0303481	.0094819	3.20	0.001	.011764	.0489323
unrate	-.0545423	.0006965	-78.31	0.000	-.0559074	-.0531772
andal	.3065884	.0099426	30.84	0.000	.2871013	.3260754
aragon	-.1747361	.0182807	-9.56	0.000	-.2105656	-.1389065
astur	-.1873592	.0212773	-8.81	0.000	-.2290619	-.1456564
balear	-.0039345	.0154917	-0.25	0.800	-.0342977	.0264286
canar	.0742636	.0135363	5.49	0.000	.0477329	.1007942
cantab	-.1447173	.0259933	-5.57	0.000	-.1956634	-.0937713
castman	-.052706	.0138269	-3.81	0.000	-.0798064	-.0256057
castleon	-.0268799	.0147625	-1.82	0.069	-.055814	.0020541
valenc	.0936022	.0099932	9.37	0.000	.0740159	.1131884
extrem	.1889583	.0215191	8.78	0.000	.1467816	.231135
galic	-.0712075	.0128358	-5.55	0.000	-.0963651	-.0460499
murcia	.0547211	.0164564	3.33	0.001	.0224671	.0869752
navarr	-.2173013	.0311463	-6.98	0.000	-.2783469	-.1562557
vasco	-.1743788	.0177102	-9.85	0.000	-.2090902	-.1396675
rioja	-.1462899	.0348503	-4.20	0.000	-.2145953	-.0779845
_cons	-2.319521	.0135216	-171.54	0.000	-2.346023	-2.293019

```
Iteration 0:   log likelihood =    -394287  (not concave)
Iteration 1:   log likelihood = -390285.16
Iteration 2:   log likelihood = -389985.45
Iteration 3:   log likelihood = -389666.04
Iteration 4:   log likelihood = -389660.59
Iteration 5:   log likelihood = -389660.57
```

```
Discrete time PH model, with discrete mixture        Number of obs   =    1547507
                                                     LR chi2()       =          .
Log likelihood = -389660.57                          Prob > chi2     =          .

------------------------------------------------------------------------------
        exit |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hazard       |
    lnjunemp |   1.667744   .0115095   144.90   0.000     1.645186    1.690302
   lnjunemp2 |  -.6094896   .0036285  -167.97   0.000    -.6166014   -.6023779
     month12 |   .2509031   .0189203    13.26   0.000     .2138199    .2879862
     month18 |   .3500861   .0319891    10.94   0.000     .2873887    .4127836
     month24 |   .8402277   .0405105    20.74   0.000     .7608285    .9196268
          ub |   -1.13203   .0224224   -50.49   0.000    -1.175977   -1.088083
 ubxlnjunemp |   .3957091   .0151461    26.13   0.000     .3660233    .4253949
      female |  -.2224591   .0087931   -25.30   0.000    -.2396933   -.2052248
     age16tv |   .1636392   .0030138    54.30   0.000     .1577322    .1695462
    age16tv2 |   -.008975   .0002028   -44.25   0.000    -.0093725   -.0085775
 educcompul1 |   .0617471   .0115434     5.35   0.000     .0391224    .0843717
 educcompul2 |   .1907854   .0088367    21.59   0.000     .1734658    .2081049
     inmigra |   .0416093   .0130047     3.20   0.001     .0161205    .0670981
      unrate |  -.0663027   .0007754   -85.51   0.000    -.0678225    -.064783
       andal |   .3799678   .0128413    29.59   0.000     .3547993    .4051363
      aragon |  -.2044564   .0254538    -8.03   0.000     -.254345   -.1545679
       astur |  -.2250878   .0304658    -7.39   0.000    -.2847996    -.165376
      balear |  -.0011708   .0222276    -0.05   0.958    -.0447362    .0423945
       canar |   .0881472   .0194367     4.54   0.000     .0500519    .1262425
      cantab |  -.1760296   .0354943    -4.96   0.000    -.2455971   -.1064621
     castman |  -.0070246   .0178225    -0.39   0.693    -.0419561    .0279069
    castleon |  -.0212195   .0199477    -1.06   0.287    -.0603162    .0178773
      valenc |   .1237376   .0136243     9.08   0.000     .0970345    .1504407
      extrem |    .268271   .0286872     9.35   0.000     .2120451    .3244969
       galic |  -.0809457    .017349    -4.67   0.000    -.1149491   -.0469424
      murcia |   .0693928   .0223672     3.10   0.002     .0255539    .1132317
      navarr |  -.2946294   .0440801    -6.68   0.000    -.3810248    -.208234
       vasco |  -.2276128   .0250431    -9.09   0.000    -.2766964   -.1785292
       rioja |  -.1576675   .0483877    -3.26   0.001    -.2525057   -.0628293
       _cons |  -2.903889   .0173715  -167.16   0.000    -2.937937   -2.869842
-------------+----------------------------------------------------------------
m2           |
       _cons |   1.037042   .0079216   130.91   0.000     1.021516    1.052568
-------------+----------------------------------------------------------------
logitp2      |
       _cons |  -.5535301   .0306487   -18.06   0.000    -.6136004   -.4934599
-------------+----------------------------------------------------------------
Prob. Type 1 |   .6349542    .007104    89.38   0.000     .6209212    .6487617
Prob. Type 2 |   .3650458    .007104    51.39   0.000     .3512383    .3790788
------------------------------------------------------------------------------
Note: m1 = 0

. display "Finished at $S_TIME"
Finished at 20:28:33
```

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

```
. display "Started at $S_TIME"
Started at 20:28:33

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

Generalized linear models                          No. of obs      =    1547507
Optimization     : ML                              Residual df     =    1547477
                                                   Scale parameter =          1
Deviance         =  787319.2776                    (1/df) Deviance =   .5087761
Pearson          =  1339880.771                    (1/df) Pearson  =   .8658486

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

                                                   AIC             =    .508805
Log likelihood   = -393659.6388                    BIC             =  -2.13e+07
```

exit	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
lnjunemp	1.529216	.0111456	137.20	0.000	1.507371	1.551061
lnjunemp2	-.6168131	.0035334	-174.57	0.000	-.6237384	-.6098879
month12	.2412328	.0188906	12.77	0.000	.2042079	.2782578
month18	.3530975	.0319817	11.04	0.000	.2904145	.4157804
month24	.8596715	.0405118	21.22	0.000	.7802699	.9390731
ub	-.9504732	.0217277	-43.74	0.000	-.9930587	-.9078876
ubxlnjunemp	.432762	.014572	29.70	0.000	.4042014	.4613226
female	-.1934465	.0061679	-31.36	0.000	-.2055354	-.1813577
age16tv	.1308036	.0025865	50.57	0.000	.1257342	.1358729
age16tv2	-.0071448	.0001769	-40.39	0.000	-.0074914	-.0067981
educcompul1	.0456268	.0081124	5.62	0.000	.0297268	.0615268
educcompul2	.1441138	.0061536	23.42	0.000	.1320529	.1561747
inmigra	.0303481	.0094819	3.20	0.001	.011764	.0489323
unrate	-.0545423	.0006965	-78.31	0.000	-.0559074	-.0531772
andal	.3065884	.0099426	30.84	0.000	.2871013	.3260754
aragon	-.1747361	.0182807	-9.56	0.000	-.2105656	-.1389065
astur	-.1873592	.0212773	-8.81	0.000	-.2290619	-.1456564
balear	-.0039345	.0154917	-0.25	0.800	-.0342977	.0264286
canar	.0742636	.0135363	5.49	0.000	.0477329	.1007942
cantab	-.1447173	.0259933	-5.57	0.000	-.1956634	-.0937713
castman	-.052706	.0138269	-3.81	0.000	-.0798064	-.0256057
castleon	-.0268799	.0147625	-1.82	0.069	-.055814	.0020541
valenc	.0936022	.0099932	9.37	0.000	.0740159	.1131884
extrem	.1889583	.0215191	8.78	0.000	.1467816	.231135
galic	-.0712075	.0128358	-5.55	0.000	-.0963651	-.0460499
murcia	.0547211	.0164564	3.33	0.001	.0224671	.0869752
navarr	-.2173013	.0311463	-6.98	0.000	-.2783469	-.1562557
vasco	-.1743788	.0177102	-9.85	0.000	-.2090902	-.1396675
rioja	-.1462899	.0348503	-4.20	0.000	-.2145953	-.0779845
_cons	-2.319521	.0135216	-171.54	0.000	-2.346023	-2.293019

```
Iteration 0:   log likelihood = -391623.26  (not concave)
Iteration 1:   log likelihood = -389180.98
Iteration 2:   log likelihood = -389080.88
Iteration 3:   log likelihood = -388981.61
Iteration 4:   log likelihood = -388976.12
Iteration 5:   log likelihood =  -388976.1
```

Discrete time PH model, with discrete mixture   Number of obs = 1547507
LR chi2() = .
Log likelihood = -388976.1   Prob > chi2 = .

exit	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
hazard						
lnjunemp	1.712728	.0117502	145.76	0.000	1.689698	1.735758
lnjunemp2	-.6112412	.0036843	-165.91	0.000	-.6184622	-.6040201
month12	.2474808	.0189257	13.08	0.000	.2103871	.2845746
month18	.349069	.0319958	10.91	0.000	.2863583	.4117797
month24	.8442376	.0405235	20.83	0.000	.764813	.9236622
ub	-1.202758	.0228459	-52.65	0.000	-1.247535	-1.15798
ubxlnjunemp	.4153211	.0153594	27.04	0.000	.3852173	.4454249
female	-.2483482	.0096691	-25.68	0.000	-.2672994	-.229397
age16tv	.1742109	.0031385	55.51	0.000	.1680596	.1803622
age16tv2	-.0096088	.0002126	-45.20	0.000	-.0100255	-.0091921
educcompul1	.0692114	.0127609	5.42	0.000	.0442004	.0942224
educcompul2	.1954493	.009653	20.25	0.000	.1765298	.2143688
inmigra	.0173987	.0141252	1.23	0.218	-.010286	.0450835
unrate	-.0692682	.0007989	-86.70	0.000	-.070834	-.0677024
andal	.3859115	.0138425	27.88	0.000	.3587807	.4130423
aragon	-.2163093	.0270087	-8.01	0.000	-.2692454	-.1633733
astur	-.2086945	.0305259	-6.84	0.000	-.2685243	-.1488648
balear	-.0051452	.023331	-0.22	0.825	-.050873	.0405827
canar	.0938902	.0203456	4.61	0.000	.0540136	.1337669
cantab	-.1922947	.0393454	-4.89	0.000	-.2694104	-.1151791
castman	-.0143631	.0195106	-0.74	0.462	-.0526032	.0238769
castleon	-.0274502	.0214405	-1.28	0.200	-.0694728	.0145724
valenc	.129262	.0150897	8.57	0.000	.0996867	.1588373
extrem	.2572255	.0300622	8.56	0.000	.1983047	.3161464
galic	-.0958032	.0194259	-4.93	0.000	-.1338771	-.0577292
murcia	.0681206	.0245183	2.78	0.005	.0200656	.1161756
navarr	-.2795309	.04251	-6.58	0.000	-.3628491	-.1962128
vasco	-.2360774	.0272906	-8.65	0.000	-.2895659	-.1825888
rioja	-.1887588	.0481213	-3.92	0.000	-.2830748	-.0944428
_cons	-2.288511	.0224326	-102.02	0.000	-2.332478	-2.244544
m2						
_cons	.984602	.0177674	55.42	0.000	.9497785	1.019425
m3						
_cons	-.8866318	.0105559	-83.99	0.000	-.9073211	-.8659425
logitp2						
_cons	-1.542246	.0543574	-28.37	0.000	-1.648784	-1.435707
logitp3						
_cons	-.0923742	.0484258	-1.91	0.056	-.1872871	.0025386
Prob. Type 1	.4704412	.0098954	47.54	0.000	.4511009	.4898707
Prob. Type 2	.1006275	.005503	18.29	0.000	.0903435	.1119381
Prob. Type 3	.4289313	.0122819	34.92	0.000	.4050452	.4531531

Note: m1 = 0

```
. display "Finished at $S_TIME"
Finished at 20:30:50
```