E C O N O M I C S   B U L L E T I N

On the estimation of hazard models with flexible baseline hazards and nonparametric unobserved heterogeneity

Michele Campolieti
University of Toronto

Abstract

I present an alternative approach for estimating hazard models that include flexible baseline hazards and nonparametric unobserved heterogeneity, which avoids the computational difficulties encountered by other researchers who have used this sort of model specification. My method places restrictions on the differences of the parameters in a flexible baseline hazard specification, which permits information from other time periods to be used to estimate the parameters of a specific time period. This method is illustrated with an empirical example.

This research was supported by the Connaught Fund Start−up and New Staff Matching grants at the University of Toronto.
Please address correspondence to: Michele Campolieti, Centre for Industrial Relations, University of Toronto, 121 St. George Street, Toronto, Ontario, M5S 2E8; email: campolie@chass.utoronto.ca.
Citation: Campolieti, Michele, (2003) "On the estimation of hazard models with flexible baseline hazards and nonparametric unobserved heterogeneity." *Economics Bulletin*, Vol. 3, No. 24 pp. 1−10
Submitted: June 28, 2003. **Accepted:** October 7, 2003.
URL: http://www.economicsbulletin.com/2003/volume3/EB−03C40001A.pdf

1 Introduction

As a practical matter, many applied researchers have reported difficulties when attempting to estimate hazard models that contain flexible baseline hazards and nonparametric unobserved heterogeneity (for example, Ham and Rea (1987), Meyer (1990), and Baker and Rea (1998)).[1,2] The source of these computational problems is not completely understood. However, these numerical obstacles can be overcome by placing more restrictions on the hazard model. Specifically, parametric assumptions about either the distribution of the unobserved heterogeneity or the shape of the baseline hazard can be imposed on the model. For example, Meyer (1990) retained the flexible baseline hazard specification in his model but assumed that the unobserved heterogeneity had a gamma distribution.[3] On the other hand, an applied research could also assume a functional form for the shape of the baseline hazard. For example, Ham and Rea (1987), using a discrete time multiperiod logit specification, specify the baseline hazard in their hazard model as a low order polynomial in duration. This assumption allowed them to fit the nonparametric unobserved heterogeneity distribution to their hazard model. The disadvantage of making these parametric assumptions is that they may not necessarily be innocuous. Consequently, a method that places weaker restrictions on the model might be more useful to empirical researchers.

In this paper, I propose an alternative approach that can be used to overcome these numerical problems. I use a discrete time hazard model that is based on a multiperiod logit specification, which has been used by, among others, Ham and Rea (1987) and Baker and Rea (1998). My model will include a nonparametric unobserved heterogeneity distribution and a flexible baseline hazard specification, with a large number of time-specific parameters. I place restrictions on the parameters of the baseline hazard, which will restrict how the slope of the baseline hazard can change. These restrictions allow information from adjacent time intervals to be used to estimate the parameters of other time intervals. This method improves the estimates of the unobserved heterogeneity distribution and does not, for the most part, have a large effect on the controls for observable heterogeneity. In addition, it is also fairly straightforward to implement and, consequently,

[1] Meyer (1990) used a grouped proportional hazards model specification. Ham and Rea (1987) and Baker and Rea (1998) used a multiperiod logit specification of the hazard model.

[2] Flexible baseline hazards are sometimes referred to as nonparametric, piece-wise constant or unrestricted baseline hazards in the literature.

[3] The unobserved heterogeneity distribution is usually chosen so that the likelihood function will have a closed form expression that is easily estimated, for example, the gamma and inverse-normal distributions.

should be a useful addition to an applied econometrician's tool kit.

2 Econometric Model

2.1 Specification of the Hazard Rate

I assume that the transition probability (i.e., the exit probability or hazard rate) between states for individual j is a discrete process with the following (logit) functional form

$$\lambda_j (t \mid \theta) = \frac{1}{1 + \exp\left(-y_j (t \mid \theta)\right)} \qquad , \tag{1}$$

where $y_j (t \mid \theta) = x_j (t)' \beta + h_j (t, \gamma) + \theta$, $x_j (t)$ is a vector of controls for observable heterogeneity, $h_j (t, \gamma)$ is the baseline hazard specification, with N_γ parameters $\left(\gamma_1, \gamma_2, ..., \gamma_{N_\gamma}\right)$, and θ denotes an unobserved heterogeneity term. In this discrete time formulation, a flexible specification of the baseline hazard is equivalent to a model with time-specific effects (Ham and Rea (1987)).[4]

Since including nonparametric unobserved heterogeneity and a flexible baseline hazard in the model presents computational problems, some additional structure must be placed on one of these components to obtain estimates of the unobserved heterogeneity distribution. My approach places this additional information on the parameters of the baseline hazard. Specifically, I restrict the differences in the parameters of the baseline hazard to lie on a polynomial of order d+1. These restrictions allow for information from some of the surrounding time intervals to be used to estimate the baseline hazard at a particular point in time. Kiefer (1990) argued that in models with a large number of interval-specific parameters it would be unlikely that the parameters from adjacent intervals would vary a great deal. This suggests that the information from some of the neighboring intervals can be combined to improve the coefficient estimates in each interval and smooth the pattern of variation across time intervals.

The restrictions imposed on the parameters of the baseline hazard are like those used in smoothness priors for distributed lag models (Shiller (1973)). Smoothness priors reflect prior information that an unknown function does not change slope quickly. These restrictions are contained in a matrix (denoted R_d), which contains binomial coefficients in its non-zero entries. The smoothness prior can be viewed as imposing stochastic restrictions on the parameters of a function (Taylor (1974)). However, in this paper the restrictions placed on the baseline hazard will be exact, unlike Campolieti (2000).

[4]For a time polynomial specification $h_j (t, \gamma) = \sum_{i=1}^{N_\gamma} \gamma_i t^i$.

For example, if $d = 1$ the restriction matrix forces the coefficients in the baseline hazard to lie on a second order polynomial.[5] The principal benefit of imposing exact restrictions is that it permits estimation with maximum likelihood and, consequently, makes the model easier to implement for applied researchers.

The restriction matrix R_d will be of dimension $\mathrm{N}_\gamma \times \mathrm{N}_\gamma$ and have ijth element

$$\mathrm{R}_d^{(ij)} = \begin{cases} 0 & \text{if } j - i > d + 1 \\ (-1)^{j+d+1-i} \binom{d+1}{j+d+1-i} & \text{otherwise} \end{cases} . \tag{2}$$

The non-zero elements of the R_d matrix are the coefficients of a polynomial of order $d + 1$. For example, if $\mathrm{N}_\gamma = 5$ and $d = 1$ the R_d matrix will have the following form

$$\mathrm{R}_1 = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \tag{3}$$

and the baseline hazard, denoted $h(t, \gamma, d)$, will take the form $(\gamma_1 - 2\gamma_2 + \gamma_3, \gamma_2 - 2\gamma_3 + \gamma_4, \gamma_3 - 2\gamma_4 + \gamma_5, \gamma_4 - 2\gamma_5, \gamma_5)$.

2.2 Estimation of Hazard Model

The exit probability in equation (1) is used to form the likelihood function of completed as well as censored or incomplete spells. Following Heckman and Singer (1984), I assume that the unobserved heterogeneity terms are drawn from a discrete distribution with N_θ points of support $\theta_1, \theta_2, ..., \theta_{N_\theta}$ and associated probabilities $p_1, p_2, ..., p_{N_\theta}$, where $p_{N_\theta} = 1 - \sum_{i=1}^{N_\theta - 1} p_i$.[6]

The density function for the completed spells can be written as

$$f_j^*(T_j) = \sum_{i=1}^{N_\theta} p_i f_j(T_j \mid \theta_i), \tag{4}$$

[5] Using information from other time intervals to estimate the parameters of a particular time interval imposes much weaker restrictions on the baseline hazard than parametric baseline hazard specifications, which restrict the curvature of the baseline hazard in a much more global fashion. In general, as the order of differencing, d, increases, the information from more time intervals will be used to estimate the parameters of a particular time interval. This suggests that it should be easier to fit the nonparametric unobserved heterogeneity distribution to a hazard model with a flexible baseline hazard specification as the order of d increases.

[6] The sampling distribution of the NPMLE is not known, except for a few special cases (Van der Vaart (1996)).

where $f_j(T_j \mid \theta_i) = \prod_{k=1}^{T_j-1}(1-\lambda_j(k \mid \theta_i))\lambda_j(T_j \mid \theta_i), \lambda_j(t \mid \theta_i) = \frac{1}{1+\exp(-y_j(t|\theta_i))}$, $y_j(t \mid \theta_i) = x_j(t)'\beta + h_j(t,\gamma,d) + \theta_i$, $i = 1,...,N_\theta$ is the number of mass points in the unobserved heterogeneity distribution and d is the order of differencing used to restrict the parameters of the baseline hazard. The expression for the censored spells can be computed in a similar fashion

$$S_j^*(T_j) = \sum_{i=1}^{N_\theta} p_i S_j(T_j \mid \theta_i), \tag{5}$$

where $S_j(T_j \mid \theta_i) = \prod_{k=1}^{T_j}(1-\lambda_j(k \mid \theta_i))$. Combining the expressions from equations (4) and (5) the likelihood function for all the individuals in the sample can be constructed and maximized with respect to $(\theta_1,...,\theta_{N_\theta},p_1,...,p_{N_\theta},\beta,\gamma)$.

3 Empirical Illustration

I estimated my hazard models using employment duration data from New Brunswick, Canada. Baker and Rea (1998) used these data to study the effect of a change in Canadian unemployment insurance (UI) eligibility rules on employment durations. Baker and Rea (1998) were unable to fit a nonparametric unobserved heterogeneity distribution to these data when their model contained a flexible baseline hazard. There are 999 individuals in the sample with a total of 1,518 employment spells. The controls for observable heterogeneity used in this specification are listed in Table 1. Further details about the construction of the eligibility variables as well as the other explanatory variables can be found in Baker and Rea (1993,1998).

The first column of Table 2 contains the parameter estimates from a hazard model with a flexible baseline hazard, which includes week-specific dummy variables for weeks 2 to 40.[7] The second column of this table contains the estimates from a hazard model with a parametric specification of the baseline hazard, i.e., a third order time polynomial in duration. The remaining columns of Table 2 contain the estimates from the hazard models with restrictions on the parameters of the flexible baseline hazard, with $d = 1, 3, 5, 7$ and 10.

The numerical problems encountered when fitting the nonparametric unobserved heterogeneity distribution to a model with a flexible baseline specification can be seen in column (1) of Table 2. Specifically, the standard error on one of the mass points in the unobserved heterogeneity distribution was very large. In addition, most of the probability mass was also concentrated on the second mass point in the distribution. Imposing a parametric

[7] The durations were artificially censored at 40 weeks.

assumption about the shape of the baseline hazard, i.e., a third order polynomial in duration, stabilized the estimates of the unobserved heterogeneity distribution (see column (2), Table 2).[8]

The estimates of the unobserved heterogeneity distribution were much better behaved in the hazard models with restrictions on the flexible baseline hazard (see columns (3) to (7) in Table 2). In particular, the standard errors on the mass points from those specifications were smaller than the unrestricted estimates in column (1). In addition, the t-statistics on the mass points in the unobserved heterogeneity distribution tended to increase as d was increased. This suggests that the estimation of the unobserved heterogeneity distribution becomes easier as d is increased and the information from more adjacent time intervals is used to estimate the baseline hazard at a particular point.

The parameter estimates on the controls for observable heterogeneity, from the specifications with restrictions on the baseline hazard, were not very sensitive to the order of differencing. However, the estimates for the eligibility variables, i.e., EL1, EL2 and EL3, were, to varying degrees, more sensitive to the specification of the baseline hazard.[9]

The estimate of EL1, which indicates that there is an increase in the employment hazard in the week the UI eligibility criteria are satisfied, was larger when the model included a parametric specification of the baseline hazard. For small values of d, the estimates from the specifications with restrictions on the parameters of the baseline hazard were very similar to those from the specification with no restrictions on the flexible baseline hazard. However, when d was increased to 7 and 10 the parameter estimates for EL1 became more like those from the model with the parametric baseline hazard. The estimates for EL2, which captures the effect of additional entitlement on the employment hazard, from the specifications with the restricted baseline hazards (in columns (3) to (7)) varied in magnitude without any clearly defined pattern. In addition, most of the estimates for EL2 were not statistically significant, so it is difficult to determine the effects of the restrictions on this variable. The estimates on the control for EL3 were the most sensitive to the specification of the baseline hazard. This sensitivity may occur because, as Baker and Rea (1998) note, this variable may be an additional control for duration dependence. Consequently, the restrictions

[8]This was the highest order polynomial supported by the data.

[9]Baker and Rea (1998) argued that a hazard model with a time polynomial controlling for duration dependence would make it more difficult to separate the effects of the eligibility rules on the employment hazard from measurement errors, such as digit preferences or calendar effects. These measurement errors would be captured by the eligibility variables when the time polynomial is used to control for duration dependence. However, the flexible baseline hazard would be better able to accommodate these measurement errors and provider a 'cleaner' estimate of the effect of the eligibility rules.

on the baseline hazard might induce this variable to capture some of the effects that the duration dependence specification was not able to absorb.

These estimates suggest that using the differencing restrictions on the parameters of a flexible baseline hazard can improve the estimates of the unobserved heterogeneity distribution and not have a large impact on the inferences associated with the other covariates in the model. The estimates of the unobserved heterogeneity distribution tended to improve as d (i.e., the order of differencing) was increased. However, as d was increased to 7 and 10 there were slightly larger impacts on some of the controls for observable heterogeneity. In this application, it appears that a value of d = 5 would work best. However, the optimal value for the order of differencing will probably vary from data set to data set, which suggests that some exploratory analysis with alternative values of d should be undertaken.

4 Concluding Remarks

This paper presented an alternative approach that can be used to overcome some of the numerical problems encountered when estimating a hazard model with a flexible baseline hazard specification and nonparametric unobserved heterogeneity distribution. The restrictions on the parameters of the baseline hazard, imposed by the restriction matrices, allow for information from other time intervals to be used to estimate the parameters for neighboring time periods. This method improved the estimates of the unobserved heterogeneity distribution and did not, for the most part, have a large impact on the other covariates in the model.

References

[1] Baker, M. and S. Rea Jr. (1993) "Employment Spells and Unemployment Insurance Eligibility Requirements" Institute for Policy Analysis, University of Toronto, Working Paper 9309.

[2] Baker, M. and S. Rea Jr. (1998) "Employment Spells and Unemployment Insurance Eligibility Requirements" *Review of Economics and Statistics* **80**, 80-94.

[3] Campolieti, M. (2000) "Bayesian Estimation and Smoothing of the Baseline Hazard in Discrete Time Duration Models" *Review of Economics and Statistics* **82**, 685-701.

[4] Ham, J.C. and S. A. Rea Jr. (1987) "Unemployment Insurance and Male Unemployment in Canada" *Journal of Labor Economics* **5**, 325-353.

[5] Han, A. and J. A. Hausman (1990) "Flexible Parametric Estimation of Duration and Competing Risks Models" *Journal of Applied Econometrics* **5**, 1-28.

[6] Heckman, J.J. and B. Singer (1984) "A Method of Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data" *Econometrica* **52**, 271-332.

[7] Kiefer, Nicholas M. (1990) "Econometric Methods for Grouped Data," in Panel Data and Labor Market Studies by J. Hartog, G. Ridder and J. Thuewens, Eds., Amsterdam: Elsevier Science Publisher: 97-117.

[8] Meyer, B.D. (1990) "Unemployment Insurance and Unemployment Spells" *Econometrica* **58**, 757-782.

[9] Shiller, R.J. (1973) "A Distributed Lag Estimator Derived from Smoothness Priors" *Econometrica* **41**, 775-788.

[10] Taylor, W.E. (1974) "Smoothness Priors and Stochastic Prior Restrictions in Distributed Lag Estimation" *International Economic Review* **15**, 803-804.

[11] Van der Vaart A. (1996) "Efficient maximum likelihood estimation in semiparametric mixture models" *Annals of Statistics* **24**, 862–878.

Table 1: Variable Definitions

Variable Name	Variable Definition
EL1	Dummy variable that takes the value 1 for the period in which a respondent satisfies the UI eligibility requirement; 0 otherwise
EL2	Dummy variable that takes the value 1 for the periods after a respondent has satisfied the UI eligibility requirement and is still accumulating benefits; 0 otherwise
EL3	Dummy variable that takes the value 1 for the periods in which the respondent has qualified for UI at the maximum entitlement; 0 otherwise
Year	Dummy variable that takes the value 1 in weeks during 1990; 0 otherwise
Unemployment Rate	Monthly unemployment rate in the respondent's province of residence
Hourly Earnings	The respondent's average hourly earnings in a given year (1989 dollars)
Age 16-24	Dummy variable that takes the value 1 if the respondent was between 16 and 24 years in 1998; 0 otherwise
Age 25-44	Dummy variable that takes the value 1 if the respondent was between 25 and 44 years in 1998; 0 otherwise
High School	Dummy variable that takes the value 1 if the respondent graduated from high school; 0 otherwise
Post Secondary	Dummy variable that takes the value 1 if the respondent has some post secondary education; 0 otherwise
Trade Certificate	Dummy variable that takes the value 1 if the respondent has a trade certificate; 0 otherwise
University	Dummy variable that takes the value 1 if the respondent graduated from university; 0 otherwise
Past UI Receipt	Dummy variable that takes the value 1 if the respondent received UI benefits in the year preceding the employment spell; 0 otherwise
Marital Status	Dummy variable that takes the value 1 if the respondent was married; 0 otherwise
School Attendance	Dummy variable that takes the value 1 if the respondent attended school in the year of the current week; 0 otherwise
Sex	Dummy variable that takes the value 1 for females; 0 otherwise

Table 2: Parameter Estimates

			Restrictions on Flexible Baseline Hazard				
	Flexible Baseline Hazard	Parametric Baseline Hazard (Time Polynomial in Duration)	Order of Differencing d = 1	Order of Differencing d = 3	Order of Differencing d = 5	Order of Differencing d = 7	Order of Differencing d = 10
Variable Name	(1)	(2)	(3)	(4)	(5)	(6)	(7)
EL1	0.4301**	0.7912**	0.4185**	0.3846**	0.4387**	0.5424**	0.6402**
	(0.1789)	(0.1445)	(0.1717)	(0.1642)	(0.1600)	(0.1200)	(0.0513)
EL2	0.2401	0.0021	0.2072	0.1249	0.1607	0.2384**	0.2782**
	(0.1582)	(0.1401)	(0.1411)	(0.1145)	(0.1170)	(0.0861)	(0.0458)
EL3	0.7816**	0.4741**	0.6645**	0.3383**	0.1995	0.3578**	0.3526**
	(0.2345)	(0.1976)	(0.1499)	(0.1245)	(0.1418)	(0.0899)	(0.0440)
Year	-0.2012**	-0.2217**	-0.2180**	-0.2413**	-0.2415**	-0.2421**	-0.2440**
	(0.0762)	(0.0675)	(0.0692)	(0.0740)	(0.0689)	(0.0660)	(0.0423)
Unemployment	-0.1083**	-0.1520**	-0.1206**	-0.0978**	-0.1244**	-0.1148**	-0.1325**
Rate	(0.0388)	(0.0311)	(0.0355)	(0.0387)	(0.0384)	(0.0357)	(0.0283)
Hourly	-0.0014	-0.0012	-0.0013	-0.0018	-0.0011	-0.0005	-0.0013
Earnings	(0.0085)	(0.0080)	(0.0080)	(0.0091)	(0.0088)	(0.0085)	(0.0076)
[Age 45-64]							
Age 16-24	0.1116	0.1099	0.1122	0.1373	0.1146	0.1360	0.1280**
	(0.1270)	(0.1167)	(0.1163)	(0.1686)	(0.1193)	(0.0987)	(0.0459)
Age 25-44	-0.0210	-0.0212	-0.0120	-0.0339	-0.0219	-0.0142	-0.0148
	(0.1056)	(0.0951)	(0.0933)	(0.1668)	(0.0096)	(0.0840)	(0.0443)
[Did Not Complete High School]							
High School	-0.1681	-0.1710**	-0.1721**	-0.1991**	-0.1790**	-0.1946**	-0.1940**
	(0.0999)	(0.0860)	(0.0860)	(0.0963)	(0.0882)	(0.0832)	(0.0451)
Post Secondary	-0.3300**	-0.3288**	-0.3354**	-0.3972**	-0.3514**	-0.3756**	-0.3649**
	(0.0956)	(0.0872)	(0.0871)	(0.1011)	(0.0913)	(0.0808)	(0.0448)
Trade	-0.3751**	-0.3706**	-0.3775**	-0.4443**	-0.3846**	-0.4127**	-0.4271**
Certificate	(0.1946)	(0.1833)	(0.1846)	(0.2009)	(0.1854)	(0.1329)	(0.0520)
University	-0.1181	-0.1366	-0.1259	-0.1807	-0.1490	-0.1625	-0.1463**
	(0.1954)	(0.1882)	(0.1897)	(0.2088)	(0.1892)	(0.1341)	(0.0521)

Past UI	0.3924**	0.3926**	0.3941**	0.4149**	0.4080**	0.4197**	0.3949**
Receipt	(0.0887)	(0.0765)	(0.0766)	(0.0841)	(0.0800)	(0.0743)	(0.0436)
Marital Status	-0.2006**	-0.1973**	-0.2012**	-0.2178**	-0.2047**	-0.2017**	-0.1936**
	(0.0099)	(0.0852)	(0.0852)	(0.0947)	(0.0874)	(0.0802)	(0.0443)
School	0.4940**	0.4854**	0.4957**	0.5975**	0.5075**	0.5430**	0.5118**
Attendance	(0.1076)	(0.0951)	(0.0949)	(0.1155)	(0.0102)	(0.0894)	(0.0443)
Sex	0.2480**	0.2482**	0.2504**	0.2753**	0.2545**	0.2703**	0.2689**
	(0.0829)	(0.0703)	(0.0702)	(0.0783)	(0.0737)	(0.0697)	(0.0430)
θ_1	-3.7937	-1.9378**	-1.6835	3.5224	1.9870**	1.1072**	1.5080**
	(728.77)	(0.4909)	(1.9958)	(2.4294)	(0.6535)	(0.1630)	(0.0533)
p_1	8.9E-10**	0.7443**	0.1950**	0.9327**	0.6044**	0.5946**	0.6208**
	(2.0E-14)	(0.0666)	(0.0873)	(0.0020)	(0.0332)	(0.0400)	(0.0259)
θ_2	-3.3516**	-2.0319**	-1.7790	-3.7059	2.3999**	0.3020	0.3823**
	(0.5457)	(0.4580)	(1.9802)	(2.5788)	(0.6633)	(0.1622)	(0.0532)
p_2	0.9999	0.2557	0.8050	0.0673	0.3956	0.4054	0.3792
	(----)	(----)	(----)	(----)	(----)	(----)	(----)
Log-likelihood	-4442.82	-4526.78	-4447.06	-4449.83	-4453.10	-4458.15	-4465.71

Notes: Standard errors in parentheses. Excluded reference group is square brackets.
Double asterisk (**) denotes statistically significant at 5 percent level.