Using the integrated hazard to estimate a fixed-effect MPH model with endogenous censoring: theory and application*

TIEMEN WOUTERSEN, PAUL FRIJTERS, GERARD VAN DEN BERG†

September 2001

Abstract

In this paper, we show how to estimate the fixed-effect MPH model with the use of moment conditions on the integrated hazard of repeat observations and we give the asymptotic distribution of our estimator. Contrary to other methods, our method is capable of allowing for lagged-duration dependence and the presence of endogenous censoring. The finite sample properties are that estimates of parameters are reasonable in small samples, but that estimates of standard errors are best attained by bootstrapping. We then apply our technique to an Australian data set. Our estimator suggests that taking account of endogenous censoring and fixed-effects yields markably different results to the random-effects MPH model sofar popular. Like Heckman and Borjas (1980), we find little evidence of lagged duration dependence.

KEYWORDS: Hazard model, endogenous regressors, fixed effects.

1 Introduction

The analysis of the durations of events has been a subject of some attention in econometrics for almost two decades. It has been complicated by heterogeneity of the data and by the

*We thank Chris Robinson for helpful comments and suggestions. The CIBC Human Capital and Productivity Program provided financial support for the last author. All errors are ours.

†Paul Frijters, University of Melbourne, Department of Economics, Melbourne, Australia. Gerard van den Berg, Free University, Amsterdam. Tiemen Woutersen, University of Western Ontario, Department of Economics, Social Science Center, London, Ontario, N6A 5C2, Canada. Email: twouters@uwo.ca.

fact that economic agents choose regressors given their characteristics and history. Thus, the regressors depend on heterogeneity and may be endogenous as well. A further complication is that a part of the data is often censored. This paper uses a newly developed estimation technique that allows us to estimate a hazard model with fixed effects, endogenous regressors, and endogenous censoring. We apply this model to study path dependence in the employment and unemployment history of a group of Australian workers.

The essence of our approach is to exploit the fact that the integrated hazard of any duration model is in expectation always 1 at the true parameter values. This naturally suggests that moment estimators can be used to find the true parameters. In the class of Mixed Proportional Hazard (MPH) models[1], we can furthermore avoid estimating the fixed effects by noticing that the integrated hazard is always proportional in the fixed effect, which means these fixed effects drop out when one uses ratios of integrated hazards. We show that this intuitive reasoning indeed works and derive the asymptotic distribution of our estimator.

The appeal of our estimator is that our basic moment conditions can be easily altered to allow for endogenous censoring and lagged duration dependence[2]. Endogenous censoring is a problem in most practical applications because empirical panels are usually of finite length. Take for instance the case that we want to analyse individuals with at least two unemployment spells in the context of a panel of limited length. Those individuals whose second unemployment spell are censored are not random, because they are more likely to be individuals whose first spell was long. The censoring of the second spell is thereby endogenous, and failure to allow for this endogeneity will result in biased estimates.

Blanchard and Wolfers (1999) argue that lagged duration dependence, leading to hysteresis, can at least partly explain the high unemployment rate in Europe. Similarly, a key aspect of the model of Ljundqvist and Sargent (1998) which attempts to explain unemployment in Europe is that individuals during unemployment lose skills and do not immediately get these skills back once in employment. The policy interest in the issue of lagged duration dependence is therefore large. This has been reflected in a large empirical and theoretical literature on the identification and estimation of lagged duration dependence.

In the context of MPH models, the issue of lagged duration dependence, i.e. whether past unemployment causes future unemployment, was first raised by Heckman and Borjas (1980). Using data on the labor market careers of 122 individuals, they developed an estimator where regressors are orthogonal to the unobserved individual effects and noted the sensitivity to the specification of the distribution of these random effects. The extreme sensitivity of the outcomes of duration models to the specification of the random effects has been noted many times since (Heckman 1981a, 1981b; see Baker and Melino 2001 for Monte-Carlo evidence). The current consensus (e.g. Heckman 1991; Baker and Melino 2001; or Van den Berg 2001) is that single-spell-based estimates of the MPH model are largely driven by functional form assumptions on the distribution of the random effects and the effect of duration itself (often called duration dependence or the baseline hazard). This situation has led researchers to turn to multiple-spells whenever possible. Honoré (1993) gave a nonconstructive proof for a fixed effect duration estimator that allows for lagged dependence though not for endogenous censoring. Extensions of this work (e.g. Frijters 2002, or Abbring and Van den Berg 2002) have confirmed that in the multiple spell case, lagged duration dependence is indeed identified, even in the absence of other regressors.

There still are only a couple of papers that have looked at multiple spells in order to look at lagged duration dependence[3]. All of them still apply a random-effect approach however and none of them allow for endogenous censoring. Heckman and Borjas (1980) found no effect of previous unemployment spells of unemployment on current transition rates from unemployment to employment, though the results were sensitive to distributional assumptions. Using a much larger data set, Lynch (1985, 1989) allows for several parametric distributions of the random effects and finds no effects. Omori (1997), who studied the careers of 2184 young men drawn from the US National Youth Longitudinal Study, did find that longer periods without employment lead to lower future transition rates to employment. Omori allowed for a non-parametric distribution of the random effects. Blau (1994) finds inconclusive results for the effect of earlier unemployment on the transition rates to work of older men. Frijters, Lindeboom and Van den Berg (2000) use a panel data version of a nonparametric random effect estimator and found no negative effect from previous long unemployment durations in Dutch administrative data.

The paper is organized as follows. In Section 2, we set up our integrated hazard approach to estimating MPH models. In Section 3 we look at the finite sample properties of our estimator. In Section 4, we present an empirical application where we compare our estimator to conventional random-effect MPH estimators. The final section concludes.

2 The integrated hazard approach

This section shows how to derive estimators that are based on the integrated hazard. The resulting estimators are a very common maximum likelihood and a methods of moments estimator.

We frequently use the fact that, given the exogenous regressor x, the integrated hazard is a unit exponential variate (see, e.g., Lancaster 1990).

$$Z = \int_0^T \theta(s, x) ds \sim \varepsilon(1) \tag{1}$$

So the expectation of Z equals one at the truth. Some remarks about notation: stochasts are denoted by uppercase letters and their realizations by lowercase letters. Method of moment estimators give the parameter estimates as a function of the data and are, therefore, written with lowercase letters. We focus on the sample analogue of the integrated hazard and denote it by z. The expectation of the sample analogue is denoted by Ez. Using lowercase letters unifies notation and highlights the fact that the integrated hazard is not just a stochast but also a transformation of the density function. In the remainder we will often suppress the underscript to denote an individual if this does not lead to confusion.

Now, consider an MPH model with fixed-effects, a piece-wise constant baseline hazard, and lagged duration dependence. We have per individual i at least two observed spells, of which the second may be endogenously censored:

$$
\begin{aligned}
\theta_{i1}(t, x_{is}) &= f_i \lambda(t) \\
\theta_{i2}(t, x_{is}) &= f_i e^{(x_{i2}-x_{i1})\beta+\alpha_0} t_{i1}^{\alpha_1} \lambda(t) \\
\text{where } \lambda(t) &= e^{\lambda_0} = 1 \text{ for } t \leq c_{break\_0} \\
&= e^{\lambda_m} \text{ for } c_{break\_m-1} \leq t \leq c_{break\_m}
\end{aligned} \tag{2}
$$

where t denotes duration and θ denotes a hazard rate. In this formulation we have taken a standard piece-wise constant baseline hazard $\lambda(t)$ which can arbitrarily closely approach any continuous function as the number of breaking points (=M) approaches infinity. As to interpretation, α_0 can be interpreted as 'occurrence dependence'; $t_{i1}^{\alpha_1}$ denotes lagged duration dependence; f_i denotes the fixed effect; and $(x_{i2} - x_{i1})\beta$ the effect of time-varying regressors. Note that in this formulation, f_i incorporates the effect of all regressors in the first spell. The complete set of parameters is denoted by γ. The matrix $(x_{i2} - x_{i1})$ includes K regressors and is assumed to be of full rank.

As data, we observe the duration of the first spell, $t1$, the duration in-between spell 1 and 2, te, an individual-specific calendar date at which spells are censored, (c_{data}), and the duration of the second spell $t2$ if the second spell is uncensored. When $t2 > c_{2,data}$, where $c_{2,data}$ is equal to $c_{data} - te - t1$, then the second spell is censored.

We will use the notation $s_j(c; \hat{\gamma})$ for the integrated hazard of duration c in spell j using parameters $\hat{\gamma}$.

We now define an endogenous censoring point for the first spell: $c_{1,data}$ is defined by the condition that $s_{i1}(c_{1,data}) = s_{i2}(c_{2,data})$. When there is no censoring at all, i.e. $c_{2,data} = \infty$, then $c_{1,data}$ will also be equal to ∞. We now define $c_1 = \min(c_{1,data}, t1)$ and $c_2 = \min(c_{2,data}, t2)$.

We can now use the following moments:

1. $g_1(\hat{\gamma}) = \frac{1}{N}\sum_{i=1}^{N} d_{i1}d_{i2}\left(s_{i1}(c_1) - s_{i2}(c_2)\right)$ and $g_2(\gamma) = \frac{1}{N}\sum_{i=1}^{N} d_{i1}d_{i2}\left(s_{i1}^2(c_1) - s_{i2}^2(c_2)\right)$.
Here d_{i1}, d_{i2} are indicators at the maximum censoring points: $d_{i1} = (t_1 \leq c_1)$ and $d_{i1} = (t_2 \leq c_2)$. These 2 moments can be used for the estimation of the occurrence dependence

parameter.

2. $g_{2+k}(\hat{\gamma}) = \frac{1}{N}\sum_{i=1}^{N} \Delta x_{1,i}^{k} d_{i1} d_{i2} (s_{i1}(c_1) - s_{i2}(c_2))$. These K moments are used for the estimation of the regressor coefficients.

3. $g_{2+K+m}(\hat{\gamma}) = \frac{1}{N}\sum_{i=1}^{N} d_{i1} d_{i2} (g_{1m} - g_{2m})$. These M moments are monotone in λ_m and are used to estimate the parameters for duration dependence. Here, g_{1m} and g_{2m} are defined by:

$$
\begin{aligned}
Y &= (\Delta x\hat{\beta} + \hat{\alpha}_0 + \hat{\alpha}_1 \ln(t_1) > 0) \\
s_{22,m} &= (t_2 < c_{break\_m+1}) s_2(t_2) \\
s_{21,m} &= (t_1 < c_{break\_m}) s_1(t_1) \\
s_{11,m} &= (t_1 < c_{break\_m+1}) s_1(t_1) \\
s_{12,m} &= (t_2 < c_{break\_m}) s_2(t_2) \\
g_{2m} &= Y * d_{i1} d_{i2} (s_{21,m} - s_{22,m}) \\
g_{1m} &= (1 - Y) * d_{i1} d_{i2} (s_{11,m} - s_{12,m})
\end{aligned}
$$

which hence describes an artificial censoring procedure.

We define $g(\hat{\gamma}) = [g_1 \,..\, g_{2+M+K}]'$..If we now estimate $\hat{\gamma}^{\max}$ as the maximand of $Q(\hat{\gamma}) = -g(\hat{\gamma})' g(\hat{\gamma})$ there holds

$$N^{-1/2}(\gamma - \hat{\gamma}^{\max}) \rightarrow N(0, \Sigma) \qquad (3)$$

proof: see appendix.

We may note that although these asymptotics suggest fast convergence, the estimation problem itself is not trivial because $Q(\hat{\gamma})$ is not continuous in the entire possible parameter space. After trying several smoothing procedures, simulated annealing turned out to be the best method to overcome the non-linearities.

The ideas in the procedure above are not immediately intuitively clear. We therefore go through the main ideas in a series of simple examples.

2.0.1 Example 1 (exponential model).

We can use equation (1) to estimate parameters of the hazard function. To illustrate how this is done, consider almost the simplest problem. Assume that $T_1, ..., T_N$ are independent durations with hazard $\theta(T) = e^{\mu}$ so $Z = e^{\mu}T$. The integrated hazards are independent unit exponentials: $Z = e^{\mu}T \sim \varepsilon(1)$. Equating the sample analogue of the integrated hazard to one gives:

$$\frac{\sum_{i=1}^{N} e^{\mu} t_i}{N} = 1.$$

This suggests an estimate for μ,

$$\hat{\mu} = -\ln\left(\frac{\sum_{i=1}^{N} t_i}{N}\right),$$

which is indeed the maximum likelihood estimator.

2.1 Example 2 (fixed effects in an MPH).

Suppose that we observe two completed durations for N individuals and want to estimate an exponential hazard model. Since we have more than one observation for each individual, we can allow for a fixed effect. Let x_{is} denote the vector of characteristics of individual i for spell s. If the spells are independent across individuals as well as across spells, then the hazards of individual i can be written as

$$\begin{aligned} \theta_{i1} &= v_i e^{x_{i1}\beta} \\ \theta_{i2} &= v_i e^{x_{i2}\beta} . \end{aligned}$$

A simple reparametrization gives

$$\begin{aligned} \theta_{i1} &= f_i \\ \theta_{i2} &= f_i e^{\triangle x_i \beta}, \end{aligned}$$

where $\triangle x_i = x_{i2} - x_{i1}$ and f_i an individual specific effect. Let ΔX be a matrix with the vectors $\triangle x_i$, $i = 1, .., N$, as its rows. Assume ΔX has full column rank. The integrated hazards for the first and second spell are $f_i t_{i1}$ and $f_i e^{\triangle x_i \beta} t_{i2}$, respectively. At the true parameter value, β_0, the difference between these integrated hazards equals zero in expectation. The expectation of the difference does not depend on the value of the fixed effect. Therefore

$$E(t_{i1} - e^{\triangle x_i \beta_0} t_{i2}) = 0.$$

Multiplying by the vector $\triangle x_i$ gives

$$\triangle x_i E(t_{i1} - e^{\triangle x_i \beta_0} t_{i2}) = 0.$$

The last equation suggests the following moment vector function:

$$g(\beta) = \frac{1}{N} \sum_i g_i$$

where

$$g_i(\beta) = \triangle x_i (t_{i1} - e^{\triangle x_i \beta} t_{i2}).$$

Maximizing based on the objective function $Q(\beta) = -\ g(\beta)' g(\beta)$ gives consistent estimates for β when the number of individuals goes to infinity (see Woutersen 2001).

2.2 Example 3 (lagged duration dependence):

Consider the following model of lagged duration dependence where we observe two, possibly censored, spells for each individual. The spells are independent across individuals and have the following hazards:

$$\begin{aligned} \theta_{i1} &= f_i \\ \theta_{i2} &= f_i e^{\gamma r(y_{i1}^{Data})}, \end{aligned}$$

where $r(y_{i1}^{Data}) > 0$ and either $r()$ finite or $r(y_{i1}^{Data})$ is bounded in probability. Note that in the presence of lagged duration dependence the probability of the second spell being censored depends on the length of the first spell, even if the censoring times are exogenous. We can remove the endogeneity by artificial censoring. Therefore we use $y_{i1} = \min(y_{i1}^{Data}, c_{i1})$ and $y_{i2} = \min(y_{i2}^{Data}, e^{-\gamma r(y_i)} c_{i2})$. The integrated hazards have the following form:

$$\begin{aligned} z_{i1} &= \int_0^{y_{i1}} \theta_{i1} ds = f_i y_{i1} \\ z_{i2} &= \int_0^{y_{i2}} \theta_{i2} ds = f_i e^{\gamma r(y_i)} y_{i2}. \end{aligned}$$

Note that after artificial censoring the probability of the second spell being censored does not depend on the length of the first spell anymore:

$$Prob(2^{\text{nd}} \text{ spell censored}) = E(1 - d_{i2}) = e^{-z_{\max}} \text{ where}$$
$$z_{\max} = f_i e^{\gamma r(y_i)} (e^{-\gamma r(y_i)} c_{i2}) = f_i c_{i2}.$$

After the endogeneity is removed we can use the following moment function:

$$g(\gamma) = \frac{1}{N} \sum_i \frac{\{d_{i2} z_{i1} - d_{i1} z_{i2}\}}{f_i} = \frac{1}{N} \sum_i \{d_{i2} y_{i1} - d_{i1} e^{\gamma r(y_i)} y_{i2}\}.$$

The equation $Eg(\gamma) = 0$ is uniquely solved for $\gamma = \gamma_0$ and the resulting estimator is consistent, see Woutersen (2001, theorem 6) for details.

3 Monte-Carlo Simulations

For the simulation we assume an uncensored first spell, a censored second spell at $c_{2,data} =$ 240 days, a constant baseline hazard $\lambda(t) = 1$ for any $t < break_1$, two breaking points: $break_1 = 30$ days and $break_2 = 180$ days. We also use four exogenous regressors which are random draws from independent normal distributions. We thus have $\gamma = [\alpha_0\ \alpha_1\ \beta_1\ \beta_2\ \beta_3\ \beta_4\ \lambda_1\ \lambda_2]'$

Table 1. Results for the model with endogenous censoring and two breaking points, for 10 simulations and 1000 observations

γ	Results	Mean Bias	$RMSE$	Median Bias	Med. Abs. Error
-0.1	-0.114	-0.014	0.005	-0.014	0.014
0.32	0.307	-0.013	0.004	-0.013	0.013
0.2	0.162	-0.037	0.013	-0.037	0.037
0.2	0.196	-0.003	0.001	-0.003	0.003
0.2	0.213	0.013	0.005	0.013	0.013
-0.8	-0.106	-0.026	0.009	-0.026	0.026
1	0.996	-0.004	0.001	-0.004	0.004
1	1.019	0.019	0.007	0.019	0.019

Table 2. Results for the model with endogenous censoring and two breaking points, for

100 simulations and 4000 observations

γ	Results	Mean Bias	$RMSE$	Median Bias	Med. Abs. Error
-0.1	-0.099	0.001	0.003	0.009	0.009
0.32	0.316	-0.004	0.007	0.019	0.019
0.2	0.196	-0.004	0.001	-0.003	0.003
0.2	0.201	0.001	0.014	-0.041	0.041
0.2	0.203	0.003	0.006	0.018	0.018
-0.8	-0.077	0.003	0.016	0.045	0.045
1	1.001	0.001	0.009	-0.026	0.026
1	1.002	0.002	0.009	-0.025	0.025

These simulations show that the parameter estimates have little bias for reasonably sized samples in the sense that the mean bias is less than a standard deviation of the truth.

However, the estimated variance is a lot higher than the theoretical variance (TIEMEN: IF I RECALL WELL, THIS LAST FINDING FOLLOWED FROM SOME ADDITIONAL SIMULATIONS YOU RAN WITH THE THEORETICAL AND EMPIRICAL STANDARD ERRORS).

4 An empirical application.

4.1 Data

The data are from the Australian Bureau of Statistics' Survey of Employment and Unemployment Patterns 1994-1997 (SEUP). This is the first longitudinal data set to detail the working and job-seeking experiences of the Australian population in general on a continuous basis (Australian Bureau of Statistics, 1997). Although the three-year time span of the survey is still relatively short for a panel, this length of observation period allows us to observe transitions in and out of work, particularly because those at risk of unemployment have been oversampled.

The SEUP consists of non-random samples from the population aged 15 to 59, making a total of 7572 respondents. The total sample overrepresents relatively disadvantaged group. We exclude full-time students, contributing family workers, the self-employed, and individuals with only 1 observed spell of unemployment. This left 2370 individuals.

Table 1: Summary statistics for the dependent and independent variables in the SEUP

	# obs	mean	sd. dev.
First spell of unemployment	2370	307.3	631.9
Calendar starting date		96.85	681.4
Minor disability		0.291	0.434
# of kids in household		0.739	1.058
Presence of partner		0.452	0.492
censored?	0		
Second spell of unemployment	2370	162.8	183.0
length of time between spells		148.9	180.3
Calendar starting date		548.2	295.74
Minor disability		0.298	0.438
# kids in household		0.755	1.053
Presence of partner		0.459	0.491
Censored?	673		

We constructed unemployment spells using the experiences reported on a monthly basis for the 3 years concerned, as well as retrospective information for the preceding 3 years. Unemployment is self-reported and corresponds to the ILO definition (an unemployed person is someone not working but looking for a job). Unemployment is recorded in days.

Although the data has a wealth of information, especially on time-invariant characteristics (see Frijters and Kalb 2002), we can only use time-varying characteristics. Because the amount of data we have is not huge and our primary interest is methodological, we will be restrictive in the set of variables used. We will especially avoid the use of income variables because these suffer from large amounts of measurement error and are furthermore highly endogenous. Because our primary interest is methodological and because our method explicitly deals with fixed individual traits, the non-representative nature of the sample and other shortcomings are not a major concern for the purposes of this paper.

Below a Table with the summary characteristics of the variables we use:

The main item of interest here is that the second spells are on average much shorter. A naive first-glance interpretation would be that second spells are shorter and that there may hence be no lagged duration dependence. However, about 30% of the second spells are censored, and it will especially be the case that those with long first spells will have short but censored second spells. The censoring is hence endogenous with any fixed-effect that there may exist. The data is hence an ideal candidate to see whether the integrated hazard approach can improve upon approaches that do not allow for fixed effects and the ensuing problem of endogenous censoring.

4.2 Models and results

We apply the integrated hazard model described in Section 2. The specification is the same as in the Monte-Carlo exercises. This means we use the flexible piece-wise constant baseline hazard with three distinct baseline pieces, 4 exogenous regressors, lagged duration dependence and occurrence dependence.

In order to have a comparison for our integrated hazard estimator, we also estimate a number of popular models:

1. The Proportional Hazard model (PH). This means an absence of unobserved heterogeneity. The hazards under this model are defined as:

$$
\begin{aligned}
\theta_{i1}(t, x_{is}) &= e^{x_{i1}\beta}\lambda(t) \\
\theta_{i2}(t, x_{is}) &= e^{x_{i2}\beta+\alpha_0} t_{i1}^{\alpha_1}\lambda(t)
\end{aligned}
$$

   where we specify the baseline hazard and the set of regressors analogue to the integrated hazard model.

2. The simple Mixed Proportional Hazard model (MPH 1). Here, we do allow for unobserved heterogeneity, but the unobserved component is orthogonal to observed characteristics. Hazards are defined as

$$
\begin{aligned}
\theta_{i1}(t, x_{is}) &= f_{i,1} e^{x_{i1}\beta}\lambda(t) \\
\theta_{i2}(t, x_{is}) &= f_{i,2} e^{x_{i2}\beta+\alpha_0} t_{i1}^{\alpha_1}\lambda(t)
\end{aligned}
$$

   where we again treat the baseline hazard and the choice of regressors analogue to the integrated hazard model. We approximate the unobserved heterogeneity distribution by

Table 2: Results of the integrated hazard model, the PH model and the MPH model

	Integrated Hazard model	PH	MPH 1	MPH2
Variables				
Calendar starting date	-0.202	0.643***	0.681***	0.688***
Minor disability	-0.004	-0.074***	-0.078***	-0.084***
# of kids in household	0.040	-0.086***	-0.092***	-0.095***
Presence of partner	0.047	-0.013	-0.0129	-0.013
$\lambda_1 - \lambda_0$	-0.456*	-0.300***	-0.192***	-0.221***
$\lambda_2 - \lambda_1$	-0.002	-0.073**	-0.042	-0.003
Occurrence dep.	-0.071***	0.217***	0.186***	-0.070
Lagged dur dep.	0.015	-0.169***	-0.179***	-0.113***
p_2			0.106***	0.142***
κ_2			7.019***	4.103***
Log L		-10.81068	-10.80310	-10.80152
$\chi^2(2)$ of MPH 1 versus PH			35.92	
# of observations	2370	2370	2370	2370

* Significant at 90%; ** significant at 95%; *** significant at 99%. The exogenous variables were all normaised. In all models, $c_{break\_1} = 90$, $c_{break\_2} = 270$, and $c_{break\_3} = +\infty$

a discrete-point distribution, i.e. : $p_j = P[f_{i,1} = \kappa_j] = P[f_{i,2} = \kappa_j]$ where j$\in \{1, .., J\}$. The probabilities p_j sum to one. We normalise $\kappa_1 = 1$. The likelihood is obtained by integrating out the unobserved heterogeneity distribution. In the simple specification, we assume that $f_{i,1}$ and $f_{i,2}$ are each separate draws from the unobserved heterogeneity distribution, i.e. we treat the two observations of the same individual as if they are from two different individuals.

3. The standard Mixed Proportional Hazard model (MPH 2). Here we take the same model as above, but we now make the standard assumption that $f_{i,1}$=$f_{i,2} = f_i$.

In the Table below, we give the results of each of the three methods, whereby the significance of the integrated hazard method follows from bootstrapping. The PH and the MPH models were estimated with Maximum Likelihood.

If we first compare the results of the PH-model with those of the MPH models, we see that the coefficients of all the exogenous and endogenous regressors have remained virtually the same, with the exception of occurrence dependence the baseline parameters. The latter is in line with the Monte-Carlo results of Baker and Melino (2001) who showed the sensitivity of the baseline hazard to the assumptions about the unobserved heterogeneity distribution. The likelihood ratio test between the two models shows that the inclusion of the extra 2 parameters has added significant explanatory power. The second MPH model does even better, suggesting some persistence of individual effects.

Comparing the results from the integrated hazard model with this model shows that the results change significantly when we take account of fixed effects and endogenous censoring. The duration dependence is again negative, though more pronounced than in the random-effects models and less significant. Because the unemployment system in Australia at that time had several incentives to stay unemployed for longer (see Frijters and Kalb 2002), this finding seems quite reasonable. Most of the exogenous variables change sign and loose significance under the integrated hazard model.

Most importantly, the finding of strong negative lagged duration dependence disappears with the integrated hazard method: the strong findings of positive occurrence dependence and negative lagged duration dependence found in the PH and MPH 1 disappears. These latter findings hence are clearly driven by the fact that the individuals with low transition rates are precisely the ones censored in the second spell, leading to a strong overestimate of the transition rates in the second spell and a negative relation between the length of the first and second spell. MPH 2, which assumes a constant individual effect, shows no significant occurrence dependence effect. Indeed, of the 3 comparison models, the coefficients of the MPH 2 model are closest to those of the integrated hazard model. Yet, there are still large differences in the significance and value of coefficients.

The main finding of the integrated hazard model is that occurrence dependence is significantly negative. The point estimate for lagged duration dependence is very low and insignificant, implying little effect of the length of previous unemployment, in line with the previous findings of Heckman and Borjas (1980 and Lynch 1989).

A disappointing aspect of the integrated hazard findings is that the method produced large standard errors and was quite time-consuming (estimation of the 200 bootstraps took 2 days on a 666 MhZ machine), which makes it difficult to include many parameters. Perhaps this will be overcome by using more information from the data via endogenous censoring and with faster computers.

5 Conclusions

In this paper we introduced and evaluated the use of the integrated hazard method for fixed-effect MPH models with lagged duration dependence and endogenous censoring. Monte Carlo results showed the finite sample properties to be quite good. Our empirical application shows that taking account of fixed effects and endogenous censoring changes the results substantially from those obtained from standard methods (the PH model or a random-effect MPH model). In line with the earliest papers in this field (Heckman and Borjas 1980), we find little evidence of a substantial negative effect of the length of previous unemployment spells on subsequent hazards from unemployment to employment. We do find a small negative occurrence dependence, whereby hazard rates are about 7% lower in a each additional unemployment spell. This means that expected unemployment spells are about 7% higher, holding constant all observed and fixed unobserved individual characteristics. This can be interpreted as a modest loss of job-finding skills.

Future challenges lie in incorporating the integrated hazard approach in a competing risk environment with a multitude of endogenous regressors.

6 Literature

Arellano, M., and B. E. Honoré, (2001): "Panel Data Models: Some Recent Developments," in *Handbook of Econometrics,* Vol. 5, ed. by J. Heckman and E. Leamer, Amsterdam: North-Holland.

Baltagi, B.H. (2000), *Nonstationary panels, panel cointegration, and dynamic panels*, Amsterdam JAI.

Blanchard, O. J., and J. Wolfers (1999): "The Role of Shocks and Institutions in the Rise of European Unemployment: The Aggregate Evidence," Working paper, MIT.

Dolton, P., and D. O'Neill (1996): "Unemployment Duration and the Restart Effect: Some Experimental Evidence," *The Economic Journal*, 106, 387-400.

Frijters, P., M. Lindeboom, and G. van den Berg (2000): "Persistencies in the labour market", Tinbergen Institute working paper, Amsterdam.

Heckman, J. J. (1981a): "Statistical Models for Discrete Panel Data," in *Structural Analysis of*

Discrete Data with Econometric Applications, edited by C. F. Manski and D. McFadden, pp. 114-178, MIT Press, Cambridge.

Heckman, J. J. (1981b): "The Incidental Parameter Problem and the Problem of Initial Conditions in

Estimating a Discrete Time-Discrete Dtata Stochastic Process," in *Structural Analysis of*

Discrete Data with Econometric Applications, edited by C. F. Manski and D. McFadden, pp. 179-195, MIT Press, Cambridge.

Heckman, J. J. (1991): "Identifying the Hand of the Past: Distinguishing State Dependence from Heterogeneity," *American Economic Review*, 81, 75-79.

Heckman, J. J., and G. J. Borjas (1980): "Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers for a Continuous Time Model of Heterogeneity and State Dependence," *Economica*, 47, 247-283.

Honoré, B. E. (1993): "Identification Results for Duration Models with Multiple Spells," *Review of Economic Studies*, 60, 241-246.

Horowitz, J. L. (1999): "Semiparametric Estimation of a Proportional Hazard Model with Unobserved Heterogeneity" *Econometrica*, 67, 1001-1028.

Lancaster, T. (1990): *The Econometric Analysis of Transition Data.* Cambridge: Cambridge University Press.

Ljundqvist, L. and T. J. Sargent (1998): "The European unemployment dilemma," *Journal of Policitcal Economy*, 106, 514-550.

Lynch, L. M. (1985): "State dependence in youth unemployment: a lost geneartion?," *Journal of Econometrics,* 71-84..

Lynch, L. M. (1989): "The youth labour market in the eighties: determinants of re-employment probabilities for young men and women", *Review of Economic Studies*, 37-45.

Phelps, E. (1972): *"Inflation policy and unemployment theory: the cost benefit approach to monetary planning".* London, McMillan.

Roed (1995): *"Hysteresis in unemployment", Journal of Economic Surveys, 11, 389-418.*

Van den Berg, G. J. (2001): "Duration Models: Specification, Identification, and Multiple Duration," in *Handbook of Econometrics,* Vol. 5. Amsterdam: North-Holland.

Woutersen, T. M. (2001) "Estimating the Hand of the Past: New Estimators for Hazard Models with Endogenous Regressors and Endogenous Censoring," Working paper, University of Western Ontario..

7 Appendices

Appendix 1.

Notes

[1]There are several good surveys on the large empirical and theoretical literature on the MPH model, notably Lancaster (1990) and Van den Berg (2001).

[2]The theoretical literature on lagged duration dependence (often termed hysteresis) is reviewed in e.g. Cross (1988) and Roed (1995). They review several mechanisms through which spells of unemployment negatively affect future labor market outcomes, including loss of skills, stigma, adaptation effects, and reference group effects.

[3]As to non-duration dynamic models, Baltagi (2000) and Arellano and Honoré, (2001) review recent efforts at implementing fixed-effect panel data models. As far as we know none of these adresses lagged duration dependence and endogenous censoring.