IDENTIFICATION RESULTS FOR DURATION MODELS WITH MULTIPLE SPELLS OR TIME–VARYING COVARIATES

Bo E. Honoré

Northwestern University

ABSTRACT: The main purpose of this paper is to investigate the identifiability of duration models with multiple spells. We prove that the results of Elbers and Ridder (1982) and Heckman and Singer (1984) can be generalized to multi–spell models with lagged duration dependence. We also prove that without lagged duration dependence, the identification result does not depend on moment conditions or tail conditions on the mixing distribution. This result is in contrast to Ridder's (1990) result for single–spells model. We also illustrate that Ridder's result can be overturned by time–varying covariates.

KEY WORDS: Duration Models, Identification. JEL No.: 211.

Bo E. Honoré
Department of Economics
Northwestern University
2003 Sheridan Road
Evanston, IL 60208

1. Introduction.

Elbers and Ridder (1982) and Heckman and Singer (1984) established that proportional hazard models with proportional unobserved heterogeneity of unknown distribution (mixed proportional hazard models) are identified if auxiliary assumptions on either the first moment of the mixing distribution or on the tail behavior of the mixing istribution are maintained.

This paper generalizes the results of Elbers and Ridder (1982) and Heckman and Singer (1984) to models with multiple spells. Multiple spells can occur because there are multiple observations of durations of the same kind. In labor economics, for example, there may be data on more than one spell of unemployment for each individual (Heckman and Borjas (1980)). In marketing, there may be data for more than one interpurchase time of a given product for each household (Jain and Vilcassim (1989) and Vilcassim and Jain (1991)). In these cases, it may be reasonable to assume that the hazard is the same for all spells for the same individual. In other cases, each individual may have spells that are related, but not exactly of the same kind. In demography, for example, there may be observations of the time from marriage to first birth, from first birth to second birth, from second birth to third birth, etc. (Heckman, Hotz and Walker (1985)). Under these circumstances, it may be desirable to specify different hazards for the different events. The main problem with identification of multispell models is that if the model has "lagged duration dependence", then one of the regressors (lagged duration) will be endogeneous, and the results of Elbers and Ridder (1982) and Heckman and Singer (1984) will not apply.

In a recent paper, Ridder (1990), it was shown (as a special case) that the identification results of Elbers and Ridder (1982) and Heckman and Singer (1984) depend crucially on the assumptions made on the mixing distribution. Without normalizing assumptions like the moment condition of Elbers and Ridder (1982), or the tail condition of Heckman and Singer (1984), the model is not identified. A special case of this result was also given in Heckman and Singer (1985), page 64. In the analysis of multiple spells, we find that if there is no lagged duration dependence and if the unobserved component (for a given individual) is the same for different spells, then it is not necessary to make any assumptions about the mixing distribution. This result is in contrast to Ridder's result for single spell model. In Section 3, we give another example of how modifications in the assumptions made by Ridder can overturn his "non–identification" result.

2. Multi–Spell Duration Models.

Suppose that for covariate x there is a positive random variable, T, with hazard function given by

$$h(t;x,\theta) = Z'(t)\phi(x)\theta \tag{1}$$

where Z is the integrated hazard (assumed to be differentiable), and θ is an unobserved random component which is assumed to have a distribution G which does not depend on x. Elbers and Ridder (1982) proved that the model given by (1) is identified if the following three assumptions are satisfied (Heckman and Singer (1984) prove identification under slightly weaker assumptions):

Assumption ER1. *θ is non–degenerate with distribution function G, and $E[\theta] = 1$.*

Assumption ER2. *The function Z defined on $[0,\infty)$ can be written as the integral of a non–negative function ψ.*

Assumption ER3. *The support of x, S, is an open set in $\Re^k$, and the function ϕ is defined on S and is non–negative, differentiable and non–constant on S. Furthermore, $\phi(x_\circ) = 1$ for some fixed $x_\circ \in S$.*

It is clear that a rescaling of Z, ϕ and θ in (1) can lead to the same distribution of T, so normalizations like $E[\theta] = 1$ and $\phi(x_\circ) = 1$ cannot be avoided. More surprisingly, Ridder (1990) proved that if no assumptions are made on the distribution of θ, then (1) is only identified up to power transformations of Z and ϕ (and "power–like" transformations of G).

In this section, we consider a panel data version of (1) in which there are two observations for each realization of θ. θ then corresponds to a random effect in analysis of panel data. First, consider a model without covariates: Assume that conditional on θ, T_1 and T_2 are independent with hazards

$$h_1(t) = Z_1'(t)\theta \qquad \text{and} \qquad h_2(t) = Z_2'(t)\theta.$$

With this specification of the hazard functions, the joint survivor function of T_1 and T_2 is

$$S(t_1,t_2) = \int_0^\infty e^{-\theta Z_1(t_1) - \theta Z_2(t_2)} dG(\theta) = \mathcal{L}\big(Z_1(t_1) + Z_2(t_2)\big), \tag{2}$$

where $\mathcal{L}$ is the LaPlace transform for θ.

This model is identified:

Theorem 1. *Suppose that for $i = 1, 2$, Z_i is differentiable and non–constant, then Z_1, Z_2, and G are identified except for a normalizing (scaling) constant.*

Proof: By differentiation of (2),

$$\frac{\partial S(t_1,t_2)/\partial t_2}{\partial S(t_1,t_2)/\partial t_1} = \frac{\mathcal{L}'\big(Z_1(t_1)+Z_2(t_2)\big)Z_2'(t_2)}{\mathcal{L}'\big(Z_1(t_1)+Z_2(t_2)\big)Z_1'(t_1)} = \frac{Z_2'(t_2)}{Z_1'(t_1)} \tag{3}$$

Let $k = Z_1'(t_0)^{-1}$ for some fixed t_0 for which $Z_1' > 0$. From (3), we can then get $kZ_2'(t_2)$, and by taking the ratio of (3) evaluated at (t_0, t_2) and (t, t_2) for some t_2 for which $Z_2' > 0$, we can get $k\tilde{Z}_1'(t_1)$.

By integration we get

$$kZ_1(t) + c_1 \qquad \text{and} \qquad kZ_2(t) + c_2$$

where we can determine c_1 and c_2 by $Z_1(0) = Z_2(0) = 0$. This completes the proof. ∎

Theorem 1 states that in a panel setting, it is not necessary to make any assumptions about the mixing distribution. Moreover, covariates are not necessary for identification of the model. This implies that the result in Theorem 1 can be extended to cover identifiability of a model in which the hazards are

$$h_1(t;x) = Z_1'(t;x)\theta \qquad \text{and} \qquad h_2(t;x) = Z_2'(t;x)\theta,$$

where the distribution of θ may depend on x. With this specification of the hazard functions, the joint survivor function of T_1 and T_2 is

$$S(t_1,t_2;x) = \int_0^\infty e^{-\theta Z_1(t_1;x)-\theta Z_2(t_2;x)}\, dG(\theta;x). \tag{4}$$

By applying Theorem 1 conditional on x, we see that $Z_1(t;x)$, $Z_2(t;x)$ and $G(\theta;x)$ are identified.

Notice that in Theorem 1, it is not assumed that the baseline hazard is the same in the two spells. For some applications, it would be natural to impose this additional restriction.

The crucial assumption in the preceding discussion is that θ enters multiplicatively on the hazard and that θ is the same for each of the durations. Depending on the application, the last assumption may or may not be reasonable. We will therefore next investigate the identifiability of mixed proportional hazard models for which θ is allowed to be different in the specification of T_1 and T_2:

$$S(t_1, t_2; x) = \int_{\Re^2} e^{-\theta_1 Z_1(t_1)\phi_1(x)} e^{-\theta_2 Z_2(t_2)\phi_2(x)} \, dG(\theta_1, \theta_2) \tag{5}$$

where G is the joint distribution of (θ_1, θ_2).

If the assumptions of Elbers and Ridder (1982), or Heckman and Singer (1984), are satisfied, then Z_1, Z_2, ϕ_1, ϕ_2 and the marginal distributions of G can be identified by considering the marginal distributions of (T_1, T_2). It then follows from the uniqueness of the multi-dimensional Laplace transform that G is identified as well. We have thus established:

Theorem 2. *Let the conditional distribution of (T_1, T_2) given x be given by (5). If, for $i = 1, 2$, (Θ_i, ϕ_i, Z_i) satisfies the assumptions of ER1, ER2 and ER3 then ϕ_1, ϕ_2, Z_1, Z_2 and G are uniquely identified (except for the scale normalizations discussed earlier).*

We now turn to extensions of (5) that allow the specification for the hazard in the second spell to depend on the outcome of the first spell, i.e. we write the density of (T_1, T_2) as

$$f(t_1, t_2; x) = \int_{\Re^2} \theta_1 Z'(t_1)\phi_1(x) e^{-\theta_1 Z_1(t_1)\phi_1(x)} \theta_2 Z'(t_2; t_1)\phi_2(x; t_1) e^{-\theta_2 Z_2(t_2;t_1)\phi_2(x;t_1)} \, dG(\theta_1, \theta_2) \tag{6}$$

Models like (6) have been used, for example, in Heckman and Borjas (1980) and Heckman, Hotz and Walker (1985). The dependence of ϕ_2 and Z_2 on T_1 is usually called "lagged duration dependence".

Conditional on T_1, θ_2 is not independent of x, so we can not use the results of Elbers and Ridder (1982) or Heckman and Singer (1984) to identify ϕ_2 and Z_2 from the conditional distribution of T_2 given T_1. A separate analysis is necessary. The next theorem gives conditions sufficient to guarantee identification of (6).

Theorem 3. *The functions Z_1, Z_2, ϕ_1, ϕ_2 and G in (6) are uniquely identified (except for the scale–normalizations discussed earlier) if*

1) (Z_1, ϕ_1, Θ_1) satisfies the conditions ER1, ER2 and ER3.

2) For given t_1, (Z_2, ϕ_2, Θ_2) satisfies the conditions ER1, ER2 and ER3.

and either

3a) $\theta_2 = R(\theta_1)$ *with probability 1 for some known function* R, $E[R(\theta_1)\theta_2] < \infty$, *and* $E[\theta_1] = 1$.

3b) θ_2 *is independent of* θ_1, *and* $E[\theta_1] = E[\theta_2] = 1$.

or

3c) $Z_2(t_2;t_1)$ *does not depend on* t_1 *and* $\phi_2(x;t_1) = \phi_2(x)h(t_1)$. *Also* $h(t_1) > 0$ *for all* t_1, $E[\theta_1] = 1$, $Z_2(t^*) = 1$ *and* $h(t^*) = 1$ *for some known* t^*, $E[\theta_1\theta_2] < \infty$, *and* $\mathrm{P}(\theta_1 > 0) = 1$.

If conditions 1), 2) and 3c) are met, except that $P(\theta_1 > 0) < 1$, *then the functions* Z_1, Z_2, ϕ_1, ϕ_2, *the marginal distribution of* θ_1 *and the joint distribution of* (θ_1, θ_2) *conditional on* $\theta_1 > 0$ *are all uniquely identified.*

Proof: From the marginal distribution of T_1 we can identify ϕ_1, Z_1 and the marginal distribution of θ_1. Let x, x_0 and t_1 be given. Then

$$\text{(7)} \qquad \frac{f(t_1,t_2;x)}{f(t_1,t_2;x_0)} \longrightarrow \frac{\phi_1(x)}{\phi_1(x_0)}\frac{\phi_2(x;t_1)}{\phi_2(x_0;t_1)}\frac{E[\theta_1\theta_2 e^{-\theta_1 Z_1(t_1)\phi_1(x)}]}{E[\theta_1\theta_2 e^{-\theta_1 Z_1(t_1)\phi_1(x_0)}]} \qquad \text{as } t_2 \to 0.$$

Now consider the cases 3a, 3b and 3c separately. Under 3a, (7) becomes

$$\frac{f(t_1,t_2;x)}{f(t_1,t_2;x_0)} \longrightarrow \frac{\phi_1(x)}{\phi_1(x_0)}\frac{\phi_2(x;t_1)}{\phi_2(x_0;t_1)}\frac{E[\theta_1 R(\theta_1) e^{-\theta_1 Z_1(t_1)\phi_1(x)}]}{E[\theta_1 R(\theta_1) e^{-\theta_1 Z_1(t_1)\phi_1(x_0)}]} \qquad \text{as } t_2 \to 0.$$

Since we have already identified ϕ_1, Z_1 and the marginal distribution for θ_1, ϕ_2 is identified up to a constant. To identify Z_2, consider

$$\int_0^{t_2} f(t_1,s;x)\,ds = Z'(t_1)\phi_1(x)E[\theta_1 e^{-\theta_1 Z_1(t_1)\phi_1(x)}\big(1 - e^{-R(\theta_1)Z_2(t_2;t_1)\phi_2(x;t_1)}\big)]$$

The right hand side is a known function of $Z_2(t_2;t_1)$, since we have already identified ϕ_2, Z_1, f_1 and the marginal distribution of θ_1. We can therefore solve for $Z_2(t_2;t_1)$ to identify Z_2. Identification of G then follows from the uniqueness of the multi–dimensional Laplace transform.

Under condition 3b, (7) becomes

$$\frac{f(t_1,t_2;x)}{f(t_1,t_2;x_0)} \longrightarrow \frac{\phi_1(x)}{\phi_1(x_0)}\frac{\phi_2(x;t_1)}{\phi_2(x_0;t_1)}\frac{E[\theta_2]E[\theta_1 e^{-\theta_1 Z_1(t_1)\phi_1(x)}]}{E[\theta_2]E[\theta_1 e^{-\theta_1 Z_1(t_1)\phi_1(x_0)}]} \qquad \text{as } t_2 \to 0$$

and again we have identification of ϕ_2 up to a constant. To identify Z_2, consider

$$\int_0^{t_2} f(t_1,s;x)\,ds = Z'(t_1)\phi_1(x)E[\theta_1 e^{-\theta_1 Z_1(t_1)\phi_1(x)}]E[1 - e^{-\theta_2 Z_2(t_2;t_1)\phi_2(x;t_1)}].$$

We can therefore solve for $E[e^{-\theta_2 Z_2(t_2;t_1)\phi_2(x;t_1)}]$, and using an argument similar to the one used in Elbers and Ridder (1982), we identify Z_2 and the marginal distribution of θ_2. Since θ_1 and θ_2 are assumed independent, this concludes the proof of identification under condition 3b.

Under condition 3c, (7) becomes

$$\frac{f(t_1,t_2;x)}{f(t_1,t_2;x_0)} \longrightarrow \frac{\phi_1(x)}{\phi_1(x_0)}\frac{\phi_2(x)}{\phi_2(x_0)}\frac{h(t_1)}{h(t_1)}\frac{E[\theta_1\theta_2 e^{-\theta_1 Z_1(t_1)\phi_1(x)}]}{E[\theta_1\theta_2 e^{-\theta_1 Z_1(t_1)\phi_1(x_0)}]} \qquad \text{as } t_2 \to 0$$

Take the limit as $t_1 \to 0$, and we get identification of ϕ_2 up to a constant.

Since ϕ_1 and Z_1 are continuous, we can define a continuous function $x(t_1)$ such that $Z_1(t_1)\phi_1(x(t_1))$ is constant on an open set A containing t^*. Let $\hat{t}_1$ and $\tilde{t}_1$ belong to A and let $\hat{x} = x(\hat{t}_1)$ and $\tilde{x} = x(\tilde{t}_1)$, then

$$\frac{f(\hat{t}_1,t_2;\hat{x})}{f(\tilde{t}_1,t_2;\tilde{x})} \longrightarrow \frac{\phi_1(\hat{x})}{\phi_1(\tilde{x})}\frac{\phi_2(\hat{x})}{\phi_2(\tilde{x})}\frac{Z_1'(\hat{t}_1)}{Z_1'(\tilde{t}_1)}\frac{h(\hat{t}_1)}{h(\tilde{t}_1)}\frac{E[\theta_1\theta_2 e^{-\theta_1 Z_1(\hat{t}_1)\phi_1(\hat{x})}]}{E[\theta_1\theta_2 e^{-\theta_1 Z_1(\tilde{t}_1)\phi_1(\tilde{x})}]} \qquad \text{as } t_2 \to 0$$

This, and the assumption that $h(t^*) = 1$, determine h on A. As above

$$\int_0^{t_2} f(t_1,s;x)\,ds = Z'(t_1)\phi_1(x)E\left[\theta_1 e^{-\theta_1 Z_1(t_1)\phi_1(x)}\left(1 - e^{\theta_2 Z_2(t_2)\phi_2(x)h(t_1)}\right)\right]$$

so we can solve for

$$E[\theta_1 e^{-\theta_1 Z_1(t_1)\phi_1(x)} e^{-\theta_2 Z_2(t_2)\phi_2(x)h(t_1)}] \tag{8}$$

Letting $t_2 = t^*$ be fixed, and having already identified Z_1, ϕ_1, ϕ_2 and h (h so far only on A), we can identify

$$K(s_1,s_2) = E[\theta_1 e^{-\theta_1 s_1} e^{-\theta_2 s_2}]$$

on an open set. Since K is real analytic, it can be extended to all of $\Re_+^2$. From knowledge of K, we can solve for Z_2 by fixing t_1 and x in (8) above ($t_1 \in A$). Likewise we can solve for h on all of $\Re_+$ by fixing $t_2 = t^*$ and fixing x.

To identify G, we integrate K with respect to s_1 to get

$$\int_{s_1}^{\infty} K(\eta,s_2)\,d\eta = E[e^{-\theta_1 s_1} e^{-\theta_2 s_2}]$$

where we have used $\mathrm{P}(\theta_1 > 0) = 1$. Hence G is identified.

If $\mathrm{P}(\theta_1 > 0) < 1$, then

$$\int_{s_1}^{\infty} K(\eta, s_2)\, d\eta = \mathrm{P}(\theta_1 > 0) E[e^{-\theta_1 s_1} e^{-\theta_2 s_2} | \theta_1 > 0]$$

and since we have already identified the marginal distribution of θ_1, we have determined the Laplace transform for the joint distribution of (θ_1, θ_2) conditional on $\theta_1 > 0$.

This completes the proof. ■

The role played by conditions 1, 2, 3a and 3b in Theorem 3 is clear, whereas condition 3c calls for some comments. First, we note that the assumptions that $E[\theta_1] = 1$, $Z_2(t^*) = 1$, $h(t^*) = 1$ and $E[\theta_1\theta_2] < \infty$ reduce to assuming $E[\theta_1] < \infty$ and $E[\theta_1\theta_2] < \infty$. The rest are normalizations.

The assumption that $Z_2(t_2; t_1)\phi_2(x; t_1) = Z_2(t_2)\phi_2(x)h(t_1)$ is strong. It is, however, general enough to cover the applications in Borjas and Heckman (1980) and Heckman, Hotz and Walker (1985).

The assumption that $\mathrm{P}(\theta_1 > 0) = 1$ is needed because $\theta_1 = 0$ will imply that $T_1 = \infty$. Since (6) does not specify the distribution of T_2 conditional on $T_1 = \infty$, we can never identify the distribution of θ_2 conditional on $\theta_1 = 0$. One possible solution is to augment (6) to include the distribution of T_2 conditional on $T_1 = \infty$. However, this solution does not seem reasonable as T_2 is usually thought of as taking place after T_1. Instead we note that it follows from the proof of Theorem 3 that everything but the distribution of θ_2 conditional on $\theta_1 = 0$ can be identified even if $\mathrm{P}(\theta_1 > 0) < 1$.

3. Time–Varying Covariates.

One of the advantages of using duration models in econometrics is that it possible to incorporate time–varying covariates in a natural way. This is not the case in truncated regression models. It is therefore of interest to think about the effect of time–varying covariates on the identifiability of duration models.

It is clear that some kinds of time–varying covariates (such as time trends) can ruin identification (see also the discussion in Heckman and Singer (1985)). The next example demonstrates that in other cases, time–varying covariates can help identification.

Consider the model given by (1). Suppose that there are two types of covariates each occurring with positive probability. The first kind of covariate is time–invariant and satisfies the conditions of Elbers and Ridder (1982) and Ridder (1990). The other kind of covariate is x_1 for $t < t^*$ and x_2 for $t \geq t^*$ where t^* is a fixed time, and where x_1 and x_2 are in the support of the time–invariant covariate and satisfy $\phi(x_1) \neq \phi(x_2)$. As discussed in the previous section, it is not possible to identify the scale of Z, ϕ and θ, so we assume that $Z(t^*) = 1$ and $\phi(x_1) = 1$.

From the observations with time–invariant covariates, it is possible identify a $\bar{Z}(t)$ and a $\bar{\phi}(x)$ such that $Z(t) = \bar{Z}(t)^\alpha$ and $\phi(x) = \bar{\phi}(x)^\alpha$ for some unknown $\alpha > 0$ (this follows from Ridder (1990)).

Now compare the observations that have time–invariant covariate x_1 to the observations that have covariate x_1 for $t < t^*$ and x_2 for $t \geq t^*$. For $t > t^*$, the survivor function for the former is $\mathcal{L}\big(Z(t)\phi(x_1)\big)$, whereas it is $\mathcal{L}\big(Z(t^*)\phi(x_1) + (Z(t) - Z(t^*))\phi(x_2)\big)$ for the latter. The ratio of the derivatives (with respect to t) of these survivor functions is

$$\frac{\mathcal{L}'\big(Z(t)\phi(x_1)\big)Z'(t)\phi(x_1)}{\mathcal{L}'\big(Z(t^*)\phi(x_1) + (Z(t) - Z(t^*))\phi(x_2)\big)Z'(t)\phi(x_2)}$$

The limit of this as $t \to t^*$ from the right is $\phi(x_1)/\phi(x_2) = \phi(x_2)^{-1}$ (here we assume that $Z' > 0$ in a neighborhood of t^*). This identifies α, and hence Z and ϕ. The distribution of θ is identified by the uniqueness of the LaPlace transform.

The usefulness of the preceding example is clearly limited by the restriction that the time–varying covariate makes a discrete jump. However, allowing the covariates to vary in a very simple way will drastically change the nature of the identifiability of proportional hazard models with a proportional unobserved component. It is our conjecture that the sensitivity of the parameter estimates of ϕ to different specifications of Z and G will depend on whether or not there are time–varying covariates.

4. Concluding Remarks.

This paper has investigated the identifiablity of multi–spell duration models. It turns out that the results of Elbers and Ridder (1982) and Heckman and Singer (1984) generalize in a natural way. It is also illustrated in this paper that in mixed proportional hazard models it is possible to get stronger identification results if there are time–varying covariates. Based on this, we speculate that the importance of controlling for unobserved heterogeneity will depend on whether or not there are time–varying covariates in the model.

ACKNOWLEDGEMENT: This paper extends results from my University of Chicago Ph. D. dissertation (Honoré (1987)). Support from NSF Grant No SES-8809352 is gratefully acknowledged.

References.

Elbers, C. and G. Ridder (1982): "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model." *Review of Economic Studies*, **49**, pp. 403–409.

Heckman, J. J. (1991): "Identifying the Hand of Past: Distinguishing State Dependence from Heterogeneity," *American Economic Review*, Papers and Proceedings, pp. 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**, pp. 247–283.

Heckman, J. J., V. J. Hotz and J. R. Walker (1985): "New Evidence on the Timing and Spacing of Births." *American Economic Review*, **72**, pp. 179–184.

Heckman, J. J. and B. Singer (1984): "The Identifiability of the Proportional Hazard Model," *Review of Economic Studies*, **60**, pp. 231–243.

Heckman, J. J. and B. Singer (1985): "Social Science Duration Analysis," in *Longitudinal Analysis of Labor Market Data* ed. by J. J. Heckman and B. Singer. New York: Cambridge University Press.

Honoré, B. E. (1987): "Identification and Estimation of Econometric Duration Models." Unpublished Ph. D. dissertation, University of Chicago.

Jain, D., and N. J. Vilcassim (1991): "Investigating Househould Purchase Timing Decisions: A Conditional Hazard Function Approach," *Marketing Science*, **10**, pp. 1-23.

Ridder, G. (1990): "The Non–Parametric Identification of Generalized Hazard Models," *Review of Economic Studies*, **57**, pp. 167–182.

Vilcassim N. J. , and D. Jain, (1991): "Modelling Purchase Timing and Brand Switching Behavior Incorporating Explanatory Variables and Unobserved Heterogeneity," *Journal of Marketing Research*, **28**, pp. 29–41.