Econometrica, Vol. 71, No. 5 (September, 2003), 1579–1589

THE SINGULARITY OF THE INFORMATION MATRIX OF THE MIXED PROPORTIONAL HAZARD MODEL

BY GEERT RIDDER AND TIEMEN M. WOUTERSEN[1]

This paper presents new identification conditions for the mixed proportional hazard model. In particular, the baseline hazard is assumed to be bounded away from 0 and ∞ near $t = 0$. These conditions ensure that the information matrix is nonsingular. The paper also presents an estimator for the mixed proportional hazard model that converges at rate $N^{-1/2}$.

KEYWORDS: Duration, semi-parametric efficiency bound, mixed proportional hazard model, duration dependence, heterogeneity, root N convergence.

1. INTRODUCTION

WE RECONSIDER THE EFFICIENCY BOUND for the semi-parametric mixed proportional hazard (MPH) model with parametric baseline hazard and regression function. This bound was first derived by Hahn (1994). One of his results is that if the baseline hazard is Weibull, the information matrix is singular, even if the model is semi-parametrically identified.[2] This implies that neither the Weibull parameter nor the regression coefficients can be estimated at a $N^{-1/2}$ rate[3] (Ishwaran (1996a) and Van der Vaart (1998, Theorem 25.32)).

Hahn's result had an impact on the use of MPH models in empirical research. The singularity of the information matrix seems to confirm the results of simulation studies (see, e.g., Baker and Melino (2000)) that suggest that it is difficult to estimate both the baseline hazard and the distribution of the random effects (or unobserved heterogeneity) with a sufficient degree of accuracy with the sample sizes that one encounters in practice. Indeed Honoré's (1990) estimator for the parameters of a semi-parametric Weibull MPH model converges at a rate slower than but arbitrarily close to $N^{-1/3}$. Ishwaran (1996b) shows that the Weibull parameter can be estimated at a rate $N^{-d/(2d+1)}$ if the moments of the unobserved heterogeneity up to $d+1$ are bounded. Altogether these results seem to imply that although the MPH model is semi-parametrically and even nonparametrically identified, the estimation of the parameters of a semi-parametric MPH model requires a larger dataset than usual.

In this note we show that this impression is false. In particular, we show that the information matrix is singular if and only if the parametric model of the integrated baseline hazard is closed under the power transformation. A set of integrated baseline hazards $\mathcal{H}$ is closed under the power transformation if $\Lambda(t) \in \mathcal{H}$ implies $\{\Lambda(t)\}^{\alpha} \in \mathcal{H}$ for every $\alpha > 0$. The Weibull baseline hazard is the most prominent member of this class

[1]We are grateful to three referees and especially to a co-editor for helpful comments. We thank Guido Imbens who urged us to write this paper. We thank Marcel Voia and Bing Zheng for their research assistance. The second author acknowledges financial support from the Social Science and Humanities Council in Canada. All errors are ours.

[2]The singularity holds if we have single-spell duration data. Hahn shows that the efficiency bound is nonsingular, if we have two or more spells for the same individual, provided that the individual random effect is the same for both spells.

[3]That is by a regular estimator sequence (for a definition see Van der Vaart (1998, p. 115)).

of models. All models that are closed under the power transformation have a baseline hazard that is either 0 or ∞ for $t = 0$, so that the restriction that the baseline hazard at 0 is bounded away from 0 and ∞ rules out closedness under the power transformation. Under this restriction the information matrix is nonsingular.

We also show that the MPH model is semi-parametrically identified if we restrict the baseline hazard near $t = 0$ to be bounded away from 0 and ∞. Hence, there are (at least) two restrictions that are sufficient for semi-parametric identification: (i) the restriction that the mean of the unobserved multiplicative random effect is finite (Elbers and Ridder (1982)[4]), and (ii) the restriction that the baseline hazard near $t = 0$ is bounded away from 0 and ∞. The first restriction does not preclude that the information matrix is singular; the second restriction does. Hence, if we impose the second restriction, there may exist estimators that are $N^{-1/2}$ consistent. Under the first restriction, the upper bound of the rate of convergence is a logarithmic rate (Ishwaran (1996a)).

Is there empirical and theoretical support for the assumption that the baseline hazard is bounded away from 0 and ∞ near $t = 0$? Let a function satisfy *boundedness* if it is bounded away from 0 and ∞ near $t = 0$. First, it should be noted that boundedness of the hazard of the duration given the covariates, i.e. ignoring unobserved heterogeneity, implies the same property of the baseline hazard in the MPH model. This makes the assumption on the baseline hazard testable (the boundedness from ∞ by testing whether one over the hazard is significantly different from 0). Second, the MPH model has been used frequently in empirical studies. Strictly, the assumption that the baseline hazard satisfies boundedness is not testable without further assumptions, because the MPH model may not be identified if this assumption does not hold. However, if the baseline hazard is specified such that its value near $t = 0$ is estimated without restrictions, e.g. by using a piecewise constant hazard, one can construct a confidence interval for that value (and its inverse). Meyer (1990, 1996) estimates such an MPH model for unemployment durations and his estimates show that the baseline hazard satisfies boundedness. The same conclusion can be drawn from Kennan's (1985) study of strike durations. He does not use an MPH, but a discrete hazard model, but the daily settlement hazards are clearly positive from the start and, although the hazard is decreasing/increasing, the hazard near $t = 0$ is not exceptionally large. From a search of the empirical literature we conclude that there is prima facie evidence that the assumption that the baseline hazard satisfies boundedness holds in most, but not all, studies.[5] Third, if we think of the MPH model as a reduced form approximation of a hazard model that is derived from economic theory, then it is important to check whether theoretical models have hazards that satisfy boundedness. We can refer to Van den Berg's (1990) study of nonstationary job search. In his model the reservation wage path is bounded and this implies that if the arrival rate of job offers satisfies boundedness, then the re-employment hazard also has that property. Blau and Robins (1986) estimate the offer arrival rate and from their estimates we conclude that it satisfies the assumption.

[4]See also Jewell (1982) and Heckman and Singer (1984) who consider an alternative identifying assumption that allows for an infinite mean, but assumes that the power transformation is fixed.

[5]References can be found on our webpages, *www-rcf.usc.edu/~ridder/* and *www.sscl.uwo.ca/economics/faculty/Woutersen.*

Yoon (1985) derives a closed form solution of the nonstationary job search model that satisfies boundedness.

The MPH model can be expressed as a transformation model with a scale normalization. Horowitz (1996) derives a semi-parametric estimator for transformation models, and Horowitz (1999) proposes an estimator of the scale parameter that, if the first three moments of the multiplicative unobserved heterogeneity are bounded, converges at a rate that is arbitrarily close to $N^{-2/5}$. We develop an estimator for the scale parameter under the assumption that the baseline hazard near $t = 0$ is constant and bounded away from 0 and ∞, but no parametric assumptions are imposed on the baseline hazard for other values of t. This estimator converges at rate $N^{-1/2}$. Combining this estimator of the scale parameter with Horowitz' (1996) estimators of the other parameters in the MPH model yields estimators for the integrated baseline hazard and the regression coefficients that converge at rate $N^{-1/2}$.

This paper is organized as follows. In Section 2 we discuss the semi-parametric MPH model and its efficiency bound as obtained by Hahn (1994). We also give an example that shows that if we change the Weibull baseline hazard slightly so that it is bounded away from 0 and ∞ at $t = 0$, then the information matrix becomes nonsingular. Section 3 contains the main result. Section 4 discusses the implications for estimation and Section 5 concludes.

2. THE SEMI-PARAMETRIC MPH MODEL: IDENTIFICATION AND EFFICIENCY BOUND

2.1. *The Semi-parametric MPH Model*

We consider the semi-parametric MPH model for the conditional distribution of T given a vector of nonconstant covariates X:

$$\theta(t|X, U; \alpha, \beta) = \lambda(t, \alpha)e^{\beta' X} e^{U} \tag{1}$$

with parametric baseline hazard $\lambda(t, \alpha)$, regression function $e^{\beta' X}$, and (α, β) in a parameter space that is an open subset of a Euclidean space. The unobserved covariates are captured by the random effect U. For example, for the Weibull model we have $\theta(t|X, U; \alpha, \beta) = \alpha t^{\alpha-1} e^{\beta' X} e^{U}$ where $\alpha > 0$. The unconditional (on U) integrated hazard at the population values of the parameters is defined as

$$S = \Lambda(T, \alpha_0)e^{\beta_0' X} \tag{2}$$

with $\Lambda(t, \alpha) = \int_0^t \lambda(s, \alpha) ds$. In Appendix 1, we show that

$$S \overset{d}{=} \frac{W}{e^{U}} \tag{3}$$

with W a standard exponential random variable that is independent of U, X and $\overset{d}{=}$ means that the random variables on both sides have the same distribution.

2.2. *Semi-parametric Identification*

Elbers and Ridder (1982) show that this MPH model is semi-parametrically identified if the following assumptions hold.

(A1) $\Lambda(t_0, \alpha_0) = 1$ *for some* $t_0 > 0$, *and* $\Lambda(\infty, \alpha_0) = \infty$.
(A2) $E(e^U) < \infty$.
(A3) *There are* x_1, x_2 *in the support of* X *with* $\beta_0' x_1 \neq \beta_0' x_2$ *and there is no constant in* X; U *and* X *are independent.*
(A4) *If* $\lambda(t, \alpha_0) = \lambda(t, \tilde{\alpha}_0)$ *for all* $t > 0$, *then* $\alpha_0 = \tilde{\alpha}_0$, *and if* $\beta_0' x = \tilde{\beta}_0' x$ *for all* x *in the support of* X, *then* $\beta_0 = \tilde{\beta}_0$.

The first part of Assumption A1 and the absence of a constant in X are normalizations. Assumption A4 ensures parametric identification of α_0, β_0.

We propose an alternative for Assumption A2.

(A2*) $0 < \lim_{t \downarrow 0} \lambda(t, \alpha_0) = \lambda(0, \alpha_0) < \infty$.

Ishwaran (1996a) shows that there exist a nonnegative random variable U_1 and a $\sigma > 0$ such that

$$\frac{W}{e^U} \overset{d}{=} \frac{W^\sigma}{e^{\sigma U_1}}.$$

If we omit covariates, the observationally equivalent MPH model has integrated baseline hazard $\Lambda(t, \alpha_0)^{1/\sigma}$ which does not satisfy A2*. Hence A2* precludes Ishwaran's construction of an observationally equivalent MPH model. Assumptions A1, A2*, A3, and A4 are sufficient for the semi-parametric identification of the MPH model.

PROPOSITION 1: *If the conditional distribution of* T *given* X *has a distribution with a* (*conditional*) *hazard as in* (1) *and if assumptions* A1, A2*, A3, *and* A4 *are satisfied*, *then* α_0, β_0 *and the distribution of* U *are identified*, *i.e. there are no observationally equivalent* $\tilde{\alpha}_0, \tilde{\beta}_0$.

PROOF: See Appendix 2.

Although both sets of conditions ensure that the semi-parametric MPH model is identified, they have different implications for the information bound of this model. In particular, with the finite mean assumption the information matrix can be singular, while with Assumption A2* this cannot be the case.

Examples of parametric models where Assumption A2* holds for all parameter values are the Gompertz baseline hazard, the rational log specification (Lancaster (1990)), and the normal hazard. See Klein and Moeschberger (1997) for a discussion of these specifications. Examples of models in which Assumption A2* is a parametric restriction are the piecewise-constant baseline hazard and the Box–Cox baseline hazard of Flinn and Heckman (1982).[6] Finally, the lognormal hazard does not satisfy A2* for all parameter values.

[6]The logarithm of this hazard model has the following form:

$$\ln\{\lambda(t, \alpha)\} = \gamma_1 \frac{t^{\lambda_1} - 1}{\lambda_1} + \gamma_2 \frac{t^{\lambda_2} - 1}{\lambda_2}$$

where $\lambda_2 > \lambda_1 \geq 0$; condition A2* holds if and only if $\lambda_1 > 0$. With this restriction the baseline hazard still can be nonmonotonic, e.g. 'bathtub' shaped.

2.3. *The Information Bound of the MPH Model*

Hahn (1994, p. 610) derives the efficient score of the MPH model using the following assumptions.

(B1) *$\lambda(t,\alpha)$ and $\Lambda(t,\alpha)$ are continuously differentiable with respect to α on an open set that contains α_0.*

(B2) *$E(X'X) < \infty$ and there exist nonnegative functions $\zeta_i(T,X)$, $i = 1, 2, 3$, such that*

$$\left|\frac{\partial \ln \lambda(T,\alpha)}{\partial \alpha}\right| \le \zeta_1(T),$$

$$\left|e^{\beta' X}\frac{\partial \Lambda(T,\alpha)}{\partial \alpha}\right| \le \zeta_2(T,X),$$

$$\left|X e^{\beta' X}\Lambda(T,\alpha)\right| \le \zeta_3(T,X),$$

with $\mathrm{E}(\zeta_1(T)^2) < \infty$, $\mathrm{E}(e^{2U}\zeta_i(T,X)^2) < \infty$, $i = 2, 3$.

The variance matrix of the efficient score at the population parameters α_0, β_0 is the information matrix. The efficient score is

$$l = \begin{bmatrix} l_\alpha \\ l_\beta \end{bmatrix} = \begin{bmatrix} a_{11} - a_{12}S \cdot \mathrm{E}[e^U|S] \\ a_2 - a_2 S \cdot \mathrm{E}[e^U|S] \end{bmatrix} \tag{4}$$

with

$$a_{11} = \frac{\partial \ln \lambda(T,\alpha)}{\partial \alpha} - \mathrm{E}\left[\frac{\partial \ln \lambda(T,\alpha)}{\partial \alpha}\middle| S\right],$$

$$a_{12} = \frac{\partial \ln \Lambda(T,\alpha)}{\partial \alpha} - \mathrm{E}\left[\frac{\partial \ln \Lambda(T,\alpha)}{\partial \alpha}\middle| S\right], \tag{5}$$

$$a_2 = X - \mathrm{E}(X|S) = X - \mathrm{E}(X);$$

see Hahn (1994, Theorem 1).[7] Without loss of generality we assume that $\mathrm{E}(X) = 0$.

For the Weibull baseline hazard $\lambda(t,\alpha) = \alpha t^{\alpha-1}$ we have

$$a_{11} = a_{12} = \ln T - \mathrm{E}(\ln T|S) \tag{6}$$

and by (2) $\ln T = (\ln S - \beta_0' X)/\alpha_0$ so that

$$a_{11} = a_{12} = -\frac{\beta_0'}{\alpha_0}X. \tag{7}$$

Substitution in (4) yields

$$l = \left(1 - S\mathrm{E}(e^U|S)\right)\begin{bmatrix} -\frac{\beta_0'}{\alpha_0}X \\ X \end{bmatrix} \tag{8}$$

[7] The efficient score is well-defined even if $\mathrm{E}(V) = \infty$; the proof is available at our webpages.

so that the distribution of the efficient score is singular at the population parameter values as is its variance matrix. This is the argument given by Hahn (1994, p. 614).

Note that this argument is not restricted to the Weibull baseline hazard. It applies to all integrated baseline hazards of the form $\lambda(t,\gamma,\alpha)=h(t,\gamma)^{\alpha}$ with h a strictly increasing function of t with $h(0,\gamma)=0$. However, a small modification of the Weibull baseline hazard gives a nonsingular information matrix. Consider the translated Weibull with integrated baseline hazard $\Lambda_{\varepsilon}(t,\alpha)=(t+\varepsilon)^{\alpha}-\varepsilon^{\alpha}$ with $\varepsilon>0$ a known constant. Note that this class of integrated baseline hazard models is not closed under the power transformation. Also the baseline hazard of this model is bounded away from 0 and ∞ if $\varepsilon>0$. A direct calculation shows that the information matrix is nonsingular.

3. NECESSARY AND SUFFICIENT CONDITIONS FOR THE SINGULARITY OF THE INFORMATION MATRIX

Our main result is Proposition 2.

PROPOSITION 2: *Under Assumptions $A1$, $A3$–$A4$, and $B1$–$B2$ the information matrix is singular if and only if the integrated baseline hazard is of the form $\Lambda(t,\alpha)=h(t)^{d(\alpha)}$ for α in some open neighborhood of α_0 with h a strictly increasing continuous function with $h(0)=0$, $h(\infty)=\infty$, and with $d(\alpha)>0$.*

PROOF: See Appendix 3.

The proof of Proposition 2 can be extended to the case of two or more parameters.[8] The baseline hazard that corresponds to $\Lambda(t,\alpha)=h(t)^{d(\alpha)}$ is

$$\lambda(t,\alpha_0)=d(\alpha_0)h(t)^{d(\alpha_0)-1}h'(t). \tag{9}$$

Note that the proposition only restricts $d(\alpha_0)$ to be positive. In particular, it can be either smaller or larger than 1. If $d(\alpha_0)<1$, then by (9) $\lim_{t\downarrow 0}\lambda(t,\alpha_0)=\infty$. If $d(\alpha_0)>1$, then $\lim_{t\downarrow 0}\lambda(t,\alpha_0)=0$; see Appendix 4 for details. Only if $d(\alpha_0)=1$, can the baseline hazard at 0 be bounded away from 0 and ∞. Hence we have the following theorem.

THEOREM: *If the assumptions for Proposition 2 hold, then $0<\lim_{t\downarrow 0}\lambda(t,\alpha_0)<\infty$ implies that the information matrix of the semi-parametric MPH model in (1) is nonsingular.*

4. IMPLICATIONS FOR ESTIMATION

A consequence of the theorem is that if we impose A2* there may exist estimators of the regression coefficients and the parameters of the integrated baseline hazard that converge at a rate $N^{-1/2}$. In this section we discuss some estimators for semi-parametric MPH models that satisfy A2*. We also develop an estimator for the case that the baseline hazard is constant near 0, but nonparametric for other values of t. In both cases the parameters are estimated at rate $N^{-1/2}$.

If the baseline hazard is specified for all $t\geq 0$, estimation starts from the observation that if we define $\Lambda(T,\alpha)\exp(\beta'X)=S(X,\alpha,\beta)$, then under weak conditions the

[8]The proof is available at our webpages.

distribution of S is independent of X if and only if $\alpha = \alpha_0$, $\beta = \beta_0$. Estimators as the quantile censoring estimator (Ridder and Woutersen (2002)) and the linear rank estimator (Bijwaard and Ridder (2002)) use this observation to formulate (potentially a continuum of) moment conditions. A proof that their moment conditions identify the parameters of the semi-parametric MPH model, even if the durations are censored, is beyond the scope of the present paper. Note that these moment conditions cannot identify the parameter σ of a power transformation of $\Lambda(t, \alpha)$ and corresponding scale of β. However, by Assumption A2* there are no observationally equivalent models with $\sigma \neq 1$.

Next consider the case that the baseline hazard is only specified near 0. Taking the logarithm of (2) gives, by Appendix 1,

$$(10) \qquad \ln \Lambda(T, \alpha) = -\beta' X - U + \ln W.$$

This is essentially a transformation model with transformation $H = \ln \Lambda$ and random error $-U + \ln W$. Horowitz (1996) suggests using existing single index estimators for β and he proposes a nonparametric estimator for H. This estimator (and the single index estimator) estimate $\ln(\Lambda(t))$ (and β) up to a multiplicative scale parameter σ. In the MPH model this scale parameter is identified either by an assumption on the moments of e^U or by Assumption A2*. Horowitz (1999) proposes an estimator for the scale parameter that converges at rate arbitrarily close to $N^{-2/5}$. Now assume, as in Meyer (1990), that the baseline hazard is constant over a small interval near 0, i.e. $0 < \lambda(t) = \lambda(0) < \infty$ for $0 \leq t \leq 2\varepsilon$. Moreover, suppose that Assumptions 1–9 of Horowitz (1996) hold and that we can estimate the transformation (up to scale) over the interval $[\varepsilon, \tau]$ where $\tau > 2\varepsilon$. Denote the estimator of the transformation by $\widehat{H(t)}$. This estimator converges at rate $N^{-1/2}$ (Horowitz (1996, Theorem 1)). Because $H(t) = \sigma \ln \Lambda(t)$, we have $H(2\varepsilon) - H(\varepsilon) = \sigma \ln 2$, so that we estimate the scale parameter σ by

$$(11) \qquad \hat{\sigma}_N = \frac{\widehat{H(2\varepsilon)} - \widehat{H(\varepsilon)}}{\ln 2}.$$

The integrated baseline hazard and the regression parameters can be estimated using $\hat{\sigma}_N$. All these estimators converge at rate $N^{-1/2}$.

5. CONCLUSION

The condition that the baseline hazard is bounded away from 0 and ∞ near $t = 0$ is sufficient for semi-parametric identification of the mixed proportional hazard model. This condition is also sufficient for a nonsingular information matrix. Hence, if the parametric baseline hazard is bounded away from 0 and ∞ near $t = 0$, there may exist (regular) estimators of the parameters of the semi-parametric MPH model with a parametric baseline hazard and regression function that are $N^{-1/2}$ consistent. In particular, we develop an estimator for the scale parameter in the MPH model (and hence the integrated baseline hazard and the regression parameters) under the assumption that the baseline hazard is constant and bounded away from 0 and ∞ in a small interval near zero. This estimator converges at rate $N^{-1/2}$.

The restriction on the baseline hazard is testable. A sufficient (but not necessary) condition for the boundedness of the baseline hazard from 0 and ∞ near $t = 0$ is that the conditional hazard given the covariates (but not the unobserved heterogeneity) and the inverse of this conditional hazard are different from 0 near $t = 0$.

Dept. of Economics, University of Southern California, Kaprielian Hall, University Park Campus, Los Angeles, CA 90089, U.S.A.; ridder@usc.edu; www-rcf.usc.edu/~ridder/

and

Dept. of Economics, University of Western Ontario, Social Science Centre, London, Ontario, N6A 5C2, Canada; twouters@uwo.ca; www.sscl.uwo.ca/economics/faculty/Woutersen.

Manuscript received November, 2001; final revision received December, 2002.

APPENDICES

APPENDIX 1: *Distribution of* $S = \Lambda(T, \alpha_0)e^{\beta_0' X}$

This distribution is derived from the relation between the (conditional on X, U) integrated hazard and the (conditional on X, U) survivor function of T. This relation is given by the product integral of the conditional integrated hazard, which is an additive interval function. A survey of the theory of product integration is given by Gill and Johansen (1990) with a useful summary in Andersen, Borgan, Gill, and Keiding (1993). If the integrated hazard exists, which in our case means that the baseline hazard is integrable, and if it is finite for finite t, then[9] by Theorem 11 of Gill and Johansen (1990), the product integral of the integrated hazard is the survivor function of a random variable T (all conditional on X, U). Moreover, because the integrated hazard is absolutely continuous, i.e. it is the integral of a function, the product integral is equal to the exponent of minus the integrated hazard, i.e. $\Pr(T > t|X, U) = \exp(-\Lambda(t, \alpha_0)e^{\beta_0' X}e^U)$. Hence $\Pr(\Lambda(T, \alpha_0)e^{\beta_0' X} > s|X, U) = \Pr(T > \Lambda^{-1}(se^{-\beta_0' X}, \alpha_0)|X, U) = e^{-se^U}$. Because U and X are independent, we have $\Pr(S > s|X) = \mathrm{E}(e^{-se^U})$ and $\Pr(S > s) = \mathrm{E}(e^{-se^U})$.

APPENDIX 2: *Proof of Proposition* 1

By (2) and (3) we have for all $t > 0$

$$\Pr(T \le t|X) = F_V\big(\Lambda(t, \alpha_0)e^{\beta_0 X}\big) \tag{12}$$

where $V = W/e^U$ is distributed as a mixture of exponential distributions and hence has a strictly increasing cdf F_V. We can assume that $\Lambda(t, \alpha_0)$ is strictly increasing in t without loss of generality. If $\tilde{\alpha}_0, \tilde{\beta}_0, \widetilde{U}$ are observationally equivalent, then for all $t > 0$

$$F_V\big(\Lambda(t, \alpha_0)e^{\beta_0 X}\big) = F_{\tilde{V}}\big(\Lambda(t, \tilde{\alpha}_0)e^{\tilde{\beta}_0 X}\big). \tag{13}$$

We denote $\Lambda(t, \alpha_0) = \Lambda(t)$, $\Lambda(t, \tilde{\alpha}_0) = \widetilde{\Lambda}(t)$, $e^{-\beta_0' x_1} = \phi_1$, $e^{-\beta_0' x_2} = \phi_2$, $e^{-\tilde{\beta}_0' x_1} = \tilde{\phi}_1$, $e^{-\tilde{\beta}_0' x_2} = \tilde{\phi}_2$ with x_1, x_2 as in A3 and without loss of generality $1 = \phi_1 > \phi_2$, $1 = \tilde{\phi}_1 > \tilde{\phi}_2$.

The inverse of a strictly increasing function exits and from (13) for all $t > 0$

$$F_V\bigg(\Lambda\big(\widetilde{\Lambda}^{-1}(t\tilde{\phi}_2)\big)\frac{1}{\phi_2}\bigg) = F_{\tilde{V}}(t) = F_V\big(\Lambda\big(\widetilde{\Lambda}^{-1}(t)\big)\big). \tag{14}$$

If we denote $K = \Lambda(\widetilde{\Lambda}^{-1}(t))$ with $K(t)$ strictly increasing and $K(0) = 0$, then (14) implies that

$$K(t\tilde{\phi}_2) = \phi_2 K(t) \tag{15}$$

[9]This assumption can be relaxed, e.g. if the end of the spell coincides with a transition to an absorbing state (see Andersen et al. (1993)). That would allow us to deal with duration distributions with a finite support. The argument can be easily extended to that case.

and by iteration for all $n \geq 1$

(16) $$K(t\tilde{\phi}_2^n) = \phi_2^n K(t).$$

If we take the derivative of (15) we obtain

(17) $$\frac{\phi_2}{\tilde{\phi}_2} K'(t) = K'(\tilde{\phi}_2 t)$$

and by iteration for all $n \geq 1$

(18) $$\left(\frac{\phi_2}{\tilde{\phi}_2}\right)^n K'(t) = K'(\phi_2^n t).$$

Taking the ratio of (18) and (16) we obtain, because

$$K'(t) = \frac{\lambda(\widetilde{\Lambda}^{-1}(t))}{\tilde{\lambda}(\widetilde{\Lambda}^{-1}(t))}$$

with $\lambda(t) = \lambda(t, \alpha_0)$, $\tilde{\lambda}(t) = \lambda(t, \tilde{\alpha}_0)$,

(19) $$\frac{K'(t)}{K(t)} = \frac{1}{t} \lim_{n\to\infty} \frac{K'(\tilde{\phi}_1^n t)}{K(\tilde{\phi}_1^n t)/\tilde{\phi}_1^n t} = \lim_{n\to\infty} \frac{1}{t} \frac{\lambda(\widetilde{\Lambda}^{-1}(\tilde{\phi}_1^n t))/\tilde{\lambda}(\widetilde{\Lambda}^{-1}(\tilde{\phi}_1^n t))}{K(\tilde{\phi}_1^n t)/\tilde{\phi}_1^n t} = \frac{1}{t}$$

by Assumption A1. Because $K(0) = 0$ this implies that $K(t) = t$ and hence $\lambda(t, \alpha_0) = \lambda(t, \tilde{\alpha}_0)$ for $t > 0$ so that $\alpha_0 = \tilde{\alpha}_0$ by A4. By (15) $\beta_0' x_2 = \tilde{\beta}_0 x_2$ for all x_2 in the support of X and hence $\beta_0 = \tilde{\beta}_0$ by A4.

APPENDIX 3: *Proof of Proposition* 2

We first rewrite the efficient score in (4) and (5) to reflect the dependence on T, X, S and the parameters,

(20) $$l = \begin{bmatrix} l_\alpha \\ l_\beta \end{bmatrix} = \begin{bmatrix} a_{11}(T, S, \alpha_0) - a_{12}(T, S, \alpha_0) H_U(S) \\ X(1 - H_U(S)) \end{bmatrix}$$

with $Z = \beta_0' X$ and $H_U(S) = S\mathrm{E}(e^U|S)$. Note that by (3) H_U does not depend on the parameters. Because S and Z are independent we have

$$a_{11}(T, S, \alpha_0) = \left.\frac{\partial \ln \lambda(T, \alpha)}{\partial \alpha}\right|_{\alpha=\alpha_0} - \mathrm{E}_Z\left[\left.\frac{\partial\{\ln \lambda(\Lambda^{-1}(Se^{-Z}, \alpha_0), \alpha)\}}{\partial \alpha}\right|_{\alpha=\alpha_0}\right],$$

$$a_{12}(T, S, \alpha_0) = \left.\frac{\partial \ln \Lambda(T, \alpha)}{\partial \alpha}\right|_{\alpha=\alpha_0} - \mathrm{E}_Z\left[\left.\frac{\partial\{\ln \Lambda(\Lambda^{-1}(Se^{-Z}, \alpha_0), \alpha)\}}{\partial \alpha}\right|_{\alpha=\alpha_0}\right],$$

where by (2) the variables T, S, Z are related by

(21) $$\ln \Lambda(T, \alpha_0) + Z = \ln S.$$

By Assumptions B1 and B2 the information bound is continuous in α_0. If the information matrix has a rank equal to the number of regressors in X, i.e. one less than full rank, for some value α_0, then by continuity it has the same rank for population parameters in a small neighborhood of α_0, $B(\alpha_0)$. Note that T depends on X only through $\beta_0' X$. By Assumption A4 the linear combination that makes the score singular must contain l_α. Because l_α depends on X only through $\beta_0' X$, loss of rank occurs if and only if l_α is proportional to $\beta_0' X$, i.e. there is a $c(\alpha) \neq 0$ on $B(\alpha_0)$ such that

(22) $$c(\alpha) a_{11}(T, S, \alpha) - c(\alpha) a_{12}(T, S, \alpha) H_U(S) = Z\{1 - H_U(S)\}$$

for α in $B(\alpha_0)$, and $S \geq 0$, Z, T that satisfy (21). From (22) it follows that for $\alpha \in B(\alpha_0)$

$$a_{11}(t, s, \alpha) = a_{12}(t, s, \alpha) \tag{23}$$

for if this equality does not hold for some $\alpha \in B(\alpha_0)$, it does not hold on some open interval, because of B1 and B2. Moreover, there is a t such that $a_{11}(t, s, \alpha)$, $a_{12}(t, s, \alpha)$ are not constant in α on that interval by Assumption A4. Hence only if the equality holds can we find a function $c(\alpha)$ such that the left-hand side does not depend on α.

Substitution in (23) gives that for all $\alpha \in B(\alpha_0)$ and $s \geq 0$ and t that satisfy (21) for some z in the support of Z

$$\frac{\partial \ln \lambda(t, \alpha)}{\partial \alpha} - \mathrm{E}_Z\left[\frac{\partial \ln \lambda(\Lambda^{-1}(se^{-Z}, \alpha), \alpha)}{\partial \alpha}\right] - \frac{\partial \ln \Lambda(t, \alpha)}{\partial \alpha} - \mathrm{E}_Z\left[\frac{\partial \ln \Lambda(\Lambda^{-1}(se^{-Z}, \alpha), \alpha)}{\partial \alpha}\right] = 0. \tag{24}$$

Note that both a_{11} and a_{12} are identically equal to 0 if Z takes only one value. If Z takes two (or more) values, then (24) holds if and only if for $\alpha \in B(\alpha_0)$ and $t > 0$

$$\frac{\partial \ln \lambda(t, \alpha)}{\partial \alpha} - \frac{\partial \ln \Lambda(t, \alpha)}{\partial \alpha} = f(\alpha). \tag{25}$$

Integrating first with respect to α and next with respect to t gives (using the initial value $\Lambda(t_0, \alpha) = 1$)

$$\ln \Lambda(t, \alpha) = e^{\int_{\alpha_0}^{\alpha} f(\gamma)d\gamma} \int_{t_0}^{t} e^{k(s)} ds \tag{26}$$

for $\alpha \in B(\alpha_0)$ and with $k(t)$ the integration constant for the integration with respect to α. Also $\int_{t_0}^{0} e^{k(s)} ds = -\infty$ and $\int_{t_0}^{\infty} e^{k(s)} ds = \infty$. If we define $h(t) = \exp(\int_{t_0}^{t} e^{k(s)} ds)$ and $d(\alpha) = \exp(\int_{\alpha_0}^{\alpha} f(\gamma)d\gamma)$, we find for $\alpha \in B(\alpha_0)$

$$\Lambda(t, \alpha) = h(t)^{d(\alpha)}$$

with h an increasing function with $h(0) = 0$ and $h(\infty) = \infty$. This completes the proof.

APPENDIX 4

This argument requires that $0 < h'(0) < \infty$. In the proof of Proposition 2 in Appendix 3, we have by (26) and because $d(\alpha_0) = 0$ that $\Lambda(t, \alpha_0) = h(t)$. Hence $0 < \lambda(0, \alpha_0) < \infty$ implies $0 < h'(0) < \infty$.

REFERENCES

ANDERSEN, P. K., O. BORGAN, R. D. GILL, AND N. KEIDING (1993): *Statistical Models Based on Counting Processes*. New York: Springer.

BAKER, M., AND A. MELINO (2000): "Duration Dependence and Nonparametric Heterogeneity: A Monte Carlo Study," *Journal of Econometrics*, 96, 357–393.

BIJWAARD, G., AND G. RIDDER (2002): "Efficient Estimation of the Semi-parametric Mixed Proportional Hazard Model," in preparation.

BLAU, D. M., AND P. K. ROBINS (1986): "Job Search, Wage Offers, and Unemployment Insurance," *Journal of Public Economics*, 29, 173–197.

ELBERS, C., AND G. RIDDER (1982): "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model," *Review of Economic Studies*, 49, 402–409.

FLINN, C. J., AND J. J. HECKMAN (1982): "Models for the Analysis of Labor Force Dynamics," in *Advances in Econometrics*, Vol. 1. Greenwich, CT: JAI Press, pp. 35–95.

GILL, R. D., AND S. JOHANSEN (1990): "A Survey of Pruduct-Integration with a View toward Application in Survival Analysis," *Annals of Statistics*, 18, 1501–1555.

HAHN, J. (1994): "The Efficiency Bound of the Mixed Proportional Hazard Model," *Review of Economic Studies*, 61, 607–629.

HECKMAN, J. J., AND B. SINGER (1984): "The Identifiability of the Proportional Hazard Model," *Review of Economic Studies*, 60, 231–243.

HONORÉ, B. E. (1990): "Simple Estimation of a Duration Model with Unobserved Heterogeneity," *Econometrica*, 58, 453–473.

HOROWITZ, J. L. (1996): "Semiparametric Estimation of the Regression Model with Unknown Transformation of the Dependent Variable," *Econometrica*, 64, 103–137.

——— (1999): "Semiparametric Estimation of a Proportional Hazard Model with Unobserved Heterogeneity," *Econometrica*, 67, 1001–1028.

ISHWARAN, H. (1996a): "Identifiability and Rates of Estimation for Scale Parameters in Location Mixture Models," *The Annals of Statistics*, 24, 1560–1571.

——— (1996b): "Uniform Rates of Estimation in the Semiparametric Weibull Mixture Model," *The Annals of Statistics*, 24, 1572–1585.

JEWELL, N. (1982): "Mixtures of Exponential Distributions," *The Annals of Statistics*, 10, 479–484.

KENNAN, J. (1985): "Contract Strikes in US Manufacturing," *Journal of Econometrics*, 28, 5–28.

KLEIN, J. P., AND M. L. MOESCHBERGER (1997): *Survival Analysis*. New York: Springer-Verlag.

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

MEYER, B. D. (1990): "Unemployment Insurance and Unemployment Spells," *Econometrica*, 58, 757–782.

——— (1996): "What Have We Learned from the Illinois Reemployment Bonus Experiment?" *Journal of Labor Economics*, 14, 26–51.

RIDDER, G. (1990): "The Non-Parametric Identification of Generalized Accelerated Failure-Time Models," *Review of Economic Studies*, 57, 167–182.

RIDDER, G., AND T. M. WOUTERSEN (2002): "Method of Moments Estimation of Duration Models with Exogenous and Endogenous Regressors," in preparation.

VAN DEN BERG, G. J. (1990): "Nonstationarity in Job Search Theory," *Review of Economic Studies*, 57, 255–277.

VAN DER VAART, A. W. (1998): *Asymptotic Statistics*. Cambridge, UK: Cambridge University Press.

YOON, B. J. (1985): "A Non-Stationary Hazard Function of Leaving Unemployment for Employment," *Economics Letters*, 17, 171–175.