Estimating a Semi-Parametric Duration Model without Specifying Heterogeneity

Jerry A. Hausman and Tiemen Woutersen[*†]
MIT and University of Arizona

First version: January 2005; this version: February 2013

Abstract. This paper presents a new estimator for the mixed proportional hazard model that allows for a nonparametric baseline hazard and time-varying regressors. In particular, this paper allows for discrete measurement of the durations as happens often in practice. The integrated baseline hazard and all parameters are estimated at the regular rate, $\sqrt{N}$, where N is the number of individuals. A hazard model is a natural framework for time-varying regressors. In particular, if a flow or a transition probability depends on a regressor that changes with time, a hazard model avoids the curse of dimensionality that would arise from interacting the regressors at each point in time with one another. This paper also presents a new test to detect unobserved heterogeneity.

Keywords: Mixed Proportional Hazard Model, Time-varying regressors, Heterogeneity.

1. Introduction

The estimation of duration models has been the subject of significant research in econometrics since the late 1970s. Since Lancaster (1979), it has been recognized that it is important to account for unobserved heterogeneity in models for duration data. Failure to account for unobserved heterogeneity causes the estimated hazard rate to decrease more with the duration than the hazard rate of a randomly selected member of the

*Correspondence addresses: Massachusetts Institute of Technology, Department of Economics, Building E52-271A, Cambridge, MA 02139; Department of Economics, Eller College of Management, University of Arizona, P.O. Box 210108, Tucson, AZ 85721; Email: jhausman@mit.edu, woutersen@email.arizona.edu

†We thank Su-Hsin Chang, Julia Driessen, Nicole Lott, Marcel Voia, and in particular Matthew Harding for research assistance. We have received helpful comments from an associate editor, two anonymous referees, Bo Honoré, Moshe Buchinsky and seminar participants at Harvard-MIT, UCLA, UCL, Texas A&M, Rice University, Yale University, UC Santa Barbara, the University of Maryland, Georgetown and the University of Virginia.

population. Moreover, the estimators of the proportional effect of explanatory variables on the population hazard rate are biased. To account for unobserved heterogeneity, Lancaster proposes a parametric Mixed Proportional Hazard (MPH) model, a generalization of Cox's (1972) Proportional Hazard model, that specifies the hazard rate as the product of a regression function that captures the effect of observed explanatory variables, a baseline hazard that captures variation in the hazard over the spell, and a random variable that accounts for the omitted heterogeneity.

Lancaster's MPH model is fully parametric, as opposed to Cox's semi-parametric approach; and from the outset, questions were raised on the role of functional form and parametric assumptions in the distinction between unobserved heterogeneity and duration dependence.[1] Elbers and Ridder (1982) resolve this question by showing that the MPH model is semi-parametrically identified if there is minimal variation in the regression function. Semi-parametric identification means that semi-parametric estimation is feasible, and a number of semi-parametric estimators for the MPH model have been proposed that progressively relaxed the parametric restrictions.

Heckman and Singer (1984) consider the nonparametric maximum likelihood estimator of the MPH model with a parametric baseline hazard and regression function. Using the results of Kiefer and Wolfowitz (1956), they approximate the unobserved heterogeneity with a discrete mixture. The rate of convergence and the asymptotic distribution of this estimator are not known. Honoré (1990) suggests another estimator that does not require specifying the unobserved heterogeneity distribution. This estimator assumes a Weibull baseline hazard and only uses very short durations to estimate the Weibull parameter.

Han and Hausman (1990) and Meyer (1990) propose an estimator that assumes that the baseline hazard is piecewise-constant, to permit flexibility, and that the heterogeneity has a gamma distribution. We present simulations and a theoretical result to show that using a nonparametric estimator of the baseline hazard with gamma heterogeneity yields inconsistent estimates for all parameters and functions if the true mixing distribution is not a gamma, which limits the usefulness of the Han-Hausman-Meyer approach. In

[1] Heckman (1991) gives an overview of attempts to make this distinction in duration and dynamic panel data models.

particular, a flexible baseline hazard does not 'compensate' for misspecification of the unobserved heterogeneity and, therefore, it is important to avoid parametric assumptions on the unobserved heterogeneity.

Horowitz (1999) was the first to propose a nonparametric estimator for both the baseline hazard and the distribution of the unobserved heterogeneity. His estimator is an adaptation of the semi-parametric estimator for a transformation model that he introduced in Horowitz (1996). In particular, if the regressors are constant over the duration, then the MPH model has a transformation model representation with the logarithm of the integrated baseline hazard as the dependent variable and a random error that is equal to the logarithm of a log standard exponential minus the logarithm of a positive random variable. In the transformation model, the regression coefficients are identified only up to a scale parameter. As shown by Ridder (1990), the scale parameter is identified in the MPH model if the unobserved heterogeneity has a finite mean. Horowitz (1999) suggests an estimator of the scale parameter that is similar to Honoré's (1990) estimator of the Weibull parameter and consistent if the finite mean assumption holds, so his approach allows estimation of the regression coefficients (not just up to scale).

The model that Horowitz (1999) estimates differs from ours. In particular, his model has regressors that do *not* change over time. Hahn (1994) shows that this model cannot be estimated at the rate $\sqrt{N}$, where N is the sample size. Ishwaran (1996a) derives the fastest possible rate at which this model can be estimated, which is $N^{2/5}$ under Horowitz's (1999) assumptions, and the rate of convergence of Horowitz's (1999) estimator is arbitrarily close to this rate. Another difference between Horowitz's (1999) estimator and ours is that Horowitz's (1999) estimator requires that the durations are measured at a continuous scale in order to estimate the transformation model. This condition often does not hold in economic data, as illustrated by the unemployment duration data that Han and Hausman (1990) discuss.[2]

In this paper, we derive a new estimator for the MPH model (with heterogeneity) that allows for a nonparametric baseline hazard and time-varying regressors. No para-

[2] Also, the estimator relies on arbitrarily short durations to estimate the scale parameter (this can be viewed as the cause of the slow convergence). Thus, the regression coefficient estimates, which are often of primary interest, are often not estimated very precisely.

metric specification of the heterogeneity distribution nor nonparametric estimation of the heterogeneity distribution is necessary. Intuitively, we condition out the heterogeneity distribution, which makes it unnecessary to estimate it. Thus, we eliminate the problems that arise with the Lancaster (1979) approach to MPH models. In our new model, the baseline hazard rate is nonparametric and the estimator of the integrated baseline hazard rate converges at the regular rate, $\sqrt{N}$, where N is the sample size. This convergence rate is the same rate as for a duration model without heterogeneity. The regressor parameters also converge at the regular rate. A nice feature of the new estimator is that it allows the durations to be measured on a finite set of points. Such discrete measurement of durations is important in economics; for example, unemployment is often measured in weeks. In the case of discrete duration measurements, the estimator of the integrated baseline hazard only converges at this set of points, as would be expected.

Bijwaard and Ridder (2002) find that the bias in the regression parameters is largely independent of the specification of the baseline hazard. Hence, failure to find significant unobserved heterogeneity should not lead to the conclusion that the bias is small.

Because it is empirically difficult to recover the distribution of the unobserved heterogeneity, estimators that rely on estimation of this distribution may be unreliable. Therefore, we avoid estimating the unobserved heterogeneity distribution.[3] Nevertheless, we can identify and estimate the regression parameters and the integrated baseline hazard. We find the removal of the requirement to estimate the heterogeneity distribution a major advantage.[4] Our estimator is related to the estimator by Han (1987). Han derives an estimator, up to scale, of the regression coefficients. However, Han's estimator cannot handle time-varying regressors, but we estimate the regression coefficients when time-varying regressors are present, as well as the scale of the regression coefficients. In particular, by estimating the regression coefficients up to scale, each regression coefficient can be interpreted as the elasticity of the hazard with respect to its regressor. Similarly, Chen's (2002) estimator of the transformation model cannot handle time-varying regressors and

[3] Horowitz (1999) also estimates his model without estimating the heterogeneity distribution and then recovers the heterogeneity distribution in a second step.

[4] An unconditional approach is also used in another context; Heckman (1978) develops unconditional tests to distinguish true and spurious state dependence.

only gives the transformation function up to scale. Also, we give an example that shows that Han's estimator is inconsistent under his assumptions and show which additional assumptions are needed.

A hazard model is a natural framework for time-varying regressors. In particular, if a flow or a transition probability depends on a regressor that changes with time, a hazard model avoids the curse of dimensionality that would arise from interacting the regressors at each point in time with one another. A nonconstructive identification proof for the duration model with time-varying regressors can be produced using techniques similar to Honoré (1993a), and Honoré (1993b) gives such a proof.[5] In particular, Honoré (1993b) does not assume that the mean of the heterogeneity distribution is finite.[6] Ridder and Woutersen (2003) argue that it is precisely the finite mean assumption that makes the identification of Elbers and Ridder (1982) 'weak' in the sense that the model of Elbers and Ridder (1982) cannot be estimated at the rate $\sqrt{N}$. As in Honoré (1993b), we do not need the finite mean assumption, which gives an intuitive explanation for why we can estimate the model at the rate $\sqrt{N}$.

This paper is organized as follows. Section 2 discusses the MPH model (with heterogeneity) and presents our estimator. Section 3 shows that our estimator converges at the regular rate and is asymptotically normally distributed. Section 4 shows that misspecifying the heterogeneity yields inconsistent estimates, even if the baseline hazard is nonparametric, and presents a new test to detect unobserved heterogeneity. Section 5 gives an empirical example and section 6 concludes.

2. Mixed Proportional Hazard Model

Lancaster (1979) introduces the MPH model in which the hazard is a function of a regressor X, unobserved heterogeneity v, and a function of time $\lambda(t)$:

$$\theta(t \mid X, v) = ve^{X\beta_0}\lambda(t). \tag{1}$$

[5] Brinch (2007) gives another nonconstructive identification proof; Woutersen (2000) and Horowitz and Lee (2004) give estimators for the panel duration model. Frederiksen, Honoré and Hu (2007) develop an estimator for a model with 'group heterogeneity,' and Honoré and Hu (2010) develop an new estimator for the transformation model. Bijwaard and Ridder (2009) derive an estimator for a MPH model with a parametric baseline hazard.

[6] Moreover, Honoré (1993b) does not assume a tail condition as in Heckman and Singer (1984).

The function $\lambda(t)$ is often referred to as the baseline hazard. The popularity of the MPH model is partly due to the fact that it nests two alternative explanations for the hazard $\theta(t|X)$ to be decreasing with time. In particular, estimating the MPH model gives the relative importance of the heterogeneity, v, and genuine duration dependence, $\lambda(t)$.[7] Lancaster (1979) uses functional form assumptions on $\lambda(t)$ and distributional assumptions on v to identify the model. Examples by Lancaster and Nickell (1980) and Heckman and Singer (1984), however, show the sensitivity to these functional form and distributional assumptions. We avoid these functional form and distributional assumptions and consider the MPH model with time-varying regressors,

$$\theta(t|X(t), v) = ve^{X(t)\beta_0}\lambda(t), \tag{2}$$

where $X(t)$ is a set of exogenous regressors whose values can vary with time, v denotes the heterogeneity, which is independent of the regressor, and $\lambda(t)$ denotes the baseline hazard. We use X to denote the sequence of the regressors that is observed for an individual. The MPH model of equation (2) implies the following survival probabilities:

$$\begin{aligned} P(T \geq t|X, v) &= \bar{F}(t|X, v) = \exp(-v\int_0^t e^{X(s)\beta_0}\lambda(s)ds) \text{ and} \\ P(T \geq t|X) &= E_v\{\bar{F}(t|X, v)\} = E_v\{\exp(-v\int_0^t e^{X(s)\beta_0}\lambda(s)ds)\}, \end{aligned} \tag{3}$$

where $E_v\{.\}$ denotes the expectation with respect to v. In applied work, durations are measured discretely; and to fix ideas, we assume that the durations are measured on a weekly scale. We also assume that the regressors can only change at the beginning of the week. Let the regressor X_{it} denote the vector of regressors for individual i during week t, i.e. $X(r) = X_s$ for $r \in [s-1, s)$. We now can write equation (3) as

$$P(T \geq t|X) = E_v\{\bar{F}(t|X, v)\} = E_v\{\exp(-v\sum_{s=1}^{t} e^{X_s\beta_0+\delta_{0,s}})\},$$

where t is a natural number, $\delta_{0,s} = \ln\{\int_{s-1}^{s}\lambda(s)dr\}$, and we normalize $\delta_{0,1} = 0$. This specification of $\delta_{0,s}$ is similar to Han-Hausman (1990), but they specify and estimate v parametrically, a requirement we remove in this paper.

Kendall (1938) proposes a statistic for rank correlation. If one is interested in the rank correlation between T and the index $X\beta$, then Kendall's (1938) rank correlation has the

[7] See Lancaster (1990) and Van den Berg (2001) for overviews.

following form:

$$Q_{\text{Kendall}}(\beta) = \frac{1}{N(N-1)} \sum_i \sum_j 1\{T_i > T_j\} 1\{X_i\beta > X_j\beta\}.$$

Han (1987) proposes an estimator that maximizes $Q_{\text{Kendall}}(\beta)$. Under certain assumptions, including that T only depends on X through the index $X\beta$, maximizing $Q_{\text{Kendall}}(\beta)$ yields an estimate for β up to scale, excluding the intercept which cannot be estimated.[8]

However, Kendall's (1938) rank correlation cannot be used for the case of time-varying regressors since it is unclear which regressor one should use. We therefore propose the following modification of the rank correlation. In particular, in our model, the expectation does depend on an index, although it has a more complicated form. Define $Z_i(l;\beta,\delta) = \sum_{s=1}^{l} e^{X_{is}\beta+\delta_s}$. We propose minimizing the following objective function:

$$Q(\beta,\delta) = \frac{1}{N(N-1)} \sum_i \sum_j \sum_{l=1}^{K} \sum_{k=1}^{K} [1\{T_i \geq l\} - 1\{T_j \geq k\}] 1\{Z_i(l;\beta,\delta) < Z_j(k;\beta,\delta)\}, \tag{4}$$

where K is the number of periods that we can observe the individuals. Thus, $Z_i(l;\beta,\delta)$ is the index *during* the l^{th} period. Intuitively, similar to Han's objective function, we compare two different individuals. However, we also take account of the outcome in each period through the parameters for the integrated hazard function, δ. The probability that individual i survives period l is larger than the probability that individual j survives period k if and only if $Z_i(l;\beta_0,\delta_0) < Z_j(k;\beta_0,\delta_0)$. The opposite holds if $Z_i(l;\beta_0,\delta_0) > Z_j(k;\beta_0,\delta_0)$. Thus, we use the outcomes for individuals i and j together with these probabilities to obtain an objective function that permits identification of the parameters β_0 and δ_0, without the restriction of only up to scale identification as in the Han approach.

Consider the expectation of the objective function, which is given by

$$E\{Q(\beta,\delta)\} = \frac{\sum_i \sum_j \sum_{l=1}^{K} \sum_{k=1}^{K}}{N(N-1)} E[\{e^{-vZ_i(l;\beta_0,\delta_0)} - e^{-vZ_j(k;\beta_0,\delta_0)}\} \cdot 1\{Z_i(l;\beta,\delta) < Z_j(k;\beta,\delta)\}].$$

This expectation is minimized at the true value of the parameters. To see this, suppose that $Z_i(l;\beta_0,\delta_0) > Z_j(k;\beta_0,\delta_0)$, so that $e^{-vZ_i(l;\beta_0,\delta_0)} < e^{-vZ_j(k;\beta_0,\delta_0)}$. Thus, $\{\beta,\delta\} =$

[8] For this reason, Han (1987) normalizes the regression coefficient by its norm, i.e. Han considers $\beta/||\beta||$. In the appendix, we present two models that satisfy Han's (1987) assumptions. The two models imply the same conditional distribution, so that one needs an additional assumption for Han's (1987) result to hold. We also present the additional assumption.

$\{\beta_0, \delta_0\}$ minimizes $E_v[\{e^{-vZ_i(l;\beta_0,\delta_0)} - e^{-vZ_j(k;\beta_0,\delta_0)}\}1\{Z_i(l;\beta,\delta) < Z_j(k;\beta,\delta)\}|X]$ for each set $\{i, j, k, l\}$ and therefore also for the expectation of the sum.[9] In contrast to the "traditional" approach that focuses on the hazard function, our approach focuses on the probability that individual i survives period l (measured from time 0). This permits a convenient treatment of the heterogeneity distribution. By only using comparisons measured from time $t = 0$, we "condition out" the heterogeneity distribution. The more traditional hazard approach considers the probability of survival conditional on individual i surviving up to period l, which requires an explicit treatment of the heterogeneity distribution.

The definition of $Q(\beta, \delta)$ given above contains a double sum, so the number of computational operations for calculating $Q(\beta, \delta)$ is N^2 (note that K is fixed). In order to reduce the number of computational operations to the order $N \ln N$, we use the rank operator. In particular, let $d_r = 1\{T \geq r\}$ for the vector T of length N. Let d be constructed by stacking the vectors d_r vertically for all $r = 1, ..., K$. Now both d and Z are vectors with length NK. If a regressor is continuously distributed conditional on the other regressors, then we can re-write $Q(\beta, \delta)$ using these vectors and the rank function, so

$$Q(\beta, \delta) = \frac{1}{N(N-1)} \sum_{j=1}^{NK} d(j)[2 \cdot Rank\{Z(j)\} - NK].$$

The computational burden to calculate[10] $Q(\beta, \delta)$ is proportional to $N \ln N$.

Note that we have identification of β rather than identification only up to an unknown scale coefficient, which is the usual outcome of most previous approaches to the problem. Also, note that by focusing on survival from the beginning of the sample, we eliminate the requirement of specifying the heterogeneity distribution since no survival bias (dynamic sample selection) occurs in our sample comparisons. Our identification is somewhat similar to the nonconstructive identification result of Elbers and Ridder (1982). However, our identification result differs in two important ways. First, our identification proof is

[9] In Appendix 1, we show that the true value *uniquely* minimizes the expectation of the objective function.

[10] Let $C(N+1)$ denote the computational cost of ordering $(N+1)$ elements given that one knows the ordering of N elements. Let m denote the median of N ordered elements and let c denote the computational cost of determining whether an element is larger than m. Then $C(2) = c$, $C(4) = 2c$, $C(2^N) = Nc$, and $C(N) \propto \ln(N)$ for any $N \geq 1$. The leading term of the computational cost of $Q(\beta, \delta)$ is proportional to $\sum_{i=1}^{N} C(i)$, which is proportional to $N \ln N$.

constructive in the sense that it suggests an estimator. Second, our identification result does *not* rely on an iterative procedure. An iterative procedure typically precludes $\sqrt{N}$ consistency.[11]

3. Large Sample Properties

In this section, we derive the large sample properties of our estimator. We assume that we observe $\{T_i, X_i\}$, where T_i is a natural number and $T_i \in [0, K]$, $K > 1$. For example, we observe unemployment duration, which is measured in weeks, and want to estimate the integrated baseline hazard at the end of each week. In order to keep the notation simple (and without loss of generality), let each period t end at $k = t$. Let $\lfloor t \rfloor$ denote the largest natural number that is smaller than or equal to t. We assume the following.

Assumption 1: *Let (i) the hazard* $\theta(t|X, v) = v \exp(X(\lfloor t \rfloor)\beta_0)\lambda(t)$, *where* $\lambda(t) \in (0, \infty)$ *for* $t \in (0, \infty)$; *(ii)* $\{T, v, X\}$ *be a random sample; (iii) the regressor* X *be exogenous, observed for* K *periods, and independent of* v; *(iv) the distribution of the regressors in the first period,* $X_{period=1}$, *not be contained in any proper linear subspace of* $\mathbb{R}^M$; *(v) the first regressor in the first period,* $X_{period=1,1}$, *has an everywhere positive density conditional on the other regressors,* $\tilde{X}_{period=1} = \tilde{x}_{period=1}$ *for almost every* $\tilde{x}_{period=1}$, *where* $\tilde{x}_{period=1} = \{x_{period=1,2}, x_{period=1,3} ..., x_{period=1,M}\}$, *i.e.* $p(x_{period=1,1}|\tilde{x}_{period=1}) > 0$ *for almost every* $\tilde{x}_{period=1}$; *(vi) the number of periods is at least two, i.e.* $K \geq 2$; *and (vii)* $\beta_0 \in \Theta$, *which is compact, and* $\beta_{0,1} \neq 0$.

The last condition, $\beta_{0,1} \neq 0$, is also necessary for Han's (1987) estimator. In particular, we give examples in the appendix that show the lack of identification under Han's (1987) assumptions. The assumptions on the distribution of the first regressor in period 1 can be relaxed at the cost of a more complicated proof.[12] Also, assumption (i) can be replaced with $Pr(T \geq t|X) = E_v \exp(-\sum_{s=1}^{s=t} v e^{X_s\beta_0 + \delta_{0,s}})$. The next assumption is about comparing the survival probabilities after the first and second period and it assumes that these are equal for some values of the regressors.

[11] Indeed, Hahn (1994) shows that the identification result of Elbers and Ridder (1982) holds for singular information matrices, so that no $\sqrt{N}$ estimator exists.

[12] See earlier versions of this paper.

ASSUMPTION 2: *Let (i)* $0 < P(T \leq 1|X_{period=1} = x_{a,1}) = P(T \leq 2|X_{period=1} = x_{b,1}, X_{period=2} = x_{b,2})$ *for some* $x_{a,1}, x_b$ *where* $x_{b,1} \neq x_{b,2}$ *and the density of the regressor is positive in an arbitrarily small neighborhood around* $x_{a,1}$ *and* $(x_{b,1}, x_{b,2})$; *(ii)* $0 < P(T \leq 1|x_{c,1}) = P(T \leq 2|x_d)$ *for some* $x_{c,1}, x_d$ *where* $x_{d,1} = x_{d,2}$ *and the density of the regressor is positive in an arbitrarily small neighborhood around* $x_{c,1}$ *and* $(x_{d,1}, x_{d,2})$

The substantial restriction in the last assumption is that a regressor changes over time, so $x_{b,1} \neq x_{b,2}$. This only has to hold after relabelling the periods. For example, one can label week 1 through week 8 as period 1 if a regressor only changes value in week 8.[13] Part (ii) of the assumption can be relaxed at the cost of a more complicated proof. Also, some of the regressors can be discrete but, as stated in the assumption, we need at least one continuously distributed regressor.

Theorem 1 (Consistency):

Let assumptions 1-2 hold. Then

$$\{\hat{\beta}, \hat{\delta}\} \underset{p}{\to} \{\beta_0, \delta_0\} \text{ and}$$

$$\sum_{s=1}^{s=t} e^{\hat{\delta}_s} \underset{p}{\to} \Lambda(t) \text{ where } t \in \{1, ..., K\}.$$

3.1 Asymptotic Distribution

In this subsection, we derive the asymptotic distribution of our estimator. As before, we use the following objective function, where $\kappa = \{\beta, \delta\}$:

$$Q(\kappa) = \frac{1}{N(N-1)} \sum_i \sum_j \sum_{l=1}^{K} \sum_{k=1}^{K} [1\{T_i \geq l\} - 1\{T_j \geq k\}] 1\{Z_i(l;\kappa) < Z_j(k;\kappa)\}.$$

In the appendix, we show that

$$Q(\kappa) = \frac{1}{N} \sum_i \sum_{l=1}^{K} 1\{T_i \geq l\} K[1 - 2\hat{F}_Z\{Z_i(l;\kappa)\}], \tag{5}$$

where $\hat{F}_Z\{Z_i(l;\kappa)\} = \frac{\sum_j}{N-1} \frac{\sum_{k=1}^{K}}{K} 1\{Z_j(k;\kappa) < Z_i(l;\kappa)\}$. Note that $\hat{F}_Z\{Z_i(l;\kappa)\}|Z_i(l;\kappa)] = F_Z\{Z_i(l;\kappa)\} + o_p(1)$, where F_Z is the cumulative distribution function of $Z_i(l;\kappa)$ for

[13] In practice, unemployment rates change every week.

$l = 1, ..., K$ and $i = 1, ..., N$. Below we assume that $Q_0(\kappa) = E\{Q(\kappa)\}$ is twice continuously differentiable at κ_0 with respect to κ. Let H denote the second derivative divided by the constant K and evaluated at κ_0, i.e.

$$H = \frac{1}{K}\nabla_{\kappa\kappa}Q_0(\kappa_0).$$

We assume the following.

ASSUMPTION 3 (INTERIOR): *Let* $\kappa_0 = (\beta_0, \delta_0) \in$ *Interior(*Θ*),* *where* Θ *is compact.*

Let $F_Z\{z(l;\kappa)\}$ denote the cumulative distribution function of $Z_i(l;\kappa)$.

ASSUMPTION 4: *Let (i)* $F_Z\{z(l;\kappa)\}$ *be twice continuously differentiable with respect to* z *in a neighborhood* $\mathcal{N}$ *of* κ_0 for any l; (ii) H *be nonsingular.*

Assumption 4 is a standard regularity condition and supports an argument based on a Taylor expansion.

Theorem 2 (Asymptotic Normality):

Let assumptions 1-4 hold. Then

$$\sqrt{N}\{\hat{\kappa} - \kappa_0\} \underset{d}{\to} N(0, H^{-1}\Omega H^{-1}),$$

where $\Omega = E[D_N(\kappa_0)D_N(\kappa_0)']$ *and*

$$D_N(\kappa) = \frac{2}{\sqrt{N}}\sum_i \sum_{l=1}^{K} [1(T_i \geq l) - E\{1(T_i \geq l)|X_i\}][\frac{\partial F_Z\{Z_i(l)\}}{\partial \kappa}|_{\kappa=\kappa_0}]$$

The function $D_N(\kappa)$ is an 'approximate derivative' and an 'influence function' in the terminology of Newey and McFadden (1994). It allows us to view the asymptotic behavior of an estimator as an average, multiplied by $\sqrt{N}$. Moreover, as Horowitz (2001, theorem 2.2) shows, bootstrapping an asymptotically normally distributed estimator that can be represented by an influence function yields a consistent variance-covariance matrix and consistent confidence intervals.[14] In the application, we bootstrap the estimator.

[14] Horowitz (2001, Theorem 2.2) averages $g_n(X_i)$.

The matrix $\Omega = E[D_N(\kappa_0)D_N(\kappa_0)']$ can be estimated using a sample analogue, where the derivative can be estimated using a kernel that omits observation i. In order to estimate H, let e_i denote the ith unit vector, let ε_N denote a small positive constant that depends on the sample size, and let $\hat{H}$ denote the matrix with (r,s)th element

$$\hat{H}_{rs} = \frac{1}{4\varepsilon_N^2}[\hat{Q}(\hat{\kappa}+e_r\varepsilon_N+e_s\varepsilon_N)-\hat{Q}(\hat{\kappa}-e_r\varepsilon_N+e_s\varepsilon_N)-\hat{Q}(\hat{\kappa}+e_r\varepsilon_N-e_s\varepsilon_N)+\hat{Q}(\hat{\kappa}-e_r\varepsilon_N-e_s\varepsilon_N)].$$

In the application, we bootstrap the estimator. In our view, choosing a value for ε_N seems arbitrary and we advise against it. See Chen (2002) for a discussion of the problems associated with choosing smoothing parameters.

Theorem 2 requires exogenous regressors. Sometimes a regressor can qualify as an exogenous regressor even if its value depends on survival up to a certain point. For example, a treatment that is randomly assigned with probability p_h to individuals who survived h periods may appear to be endogenous since it depends on survival. However, in this duration framework, we can relabel the treatment as if it is given at the beginning of the spell with probability p_h and consider the randomly assigned treatment exogenous.[15] Our estimates of $\{\delta_1, ..., \delta_K\}$ imply an estimate for the integrated hazard. In particular, if we measure survival in periods $\{0, 1, ..., K\}$, then

$$\widehat{\Lambda(0)} = 0 \text{ and } \widehat{\Lambda(t)} = \sum_{s=1}^{s=t} \exp(\hat{\delta}_s) \text{ where } t \in \{1, ..., K\}.$$

We define the average hazard on the interval $[a, b)$ as the value λ for which $\int_a^b \lambda(s)ds = \Lambda(b) - \Lambda(a)$. This gives an expression for the average hazard,

$$\widehat{\lambda(s)} = \exp(\hat{\delta}_t) \text{ for } t-1 < s \leq t.$$

If the durations are measured on a fine grid, then one could also approximate the hazard by numerically differentiating the integrated hazard $\widehat{\Lambda(t)}$. Thus, we can estimate the integrated hazard rate at each point and also approximate the hazard rate at each point. This differs considerably from Chen (2002), who only estimates the logarithm of the

[15] In particular, individuals that do not survive up to period h will be assigned treatment with probability p_h; an alternative is to use a weighting function that gives the weiths p_h and $(1-p_h)$ to both possible outcomes.

integrated hazard up to an unknown scalar, so he does not know whether the hazard is increasing or decreasing.

Another application of Theorem 2 is to compare the estimates of this paper to estimates for a more restrictive model. For example, the more restrictive model could be a model that assumes no heterogeneity (v is the same for all individuals) or assumes that v has a gamma distribution, as is a popular assumption in applied work (see, e.g., Van den Berg (2001)). If the estimator of the restrictive model is also normally distributed, then we can use the bootstrap to derive a χ^2-test. Moreover, if the estimator of the restrictive model is efficient, then we have a nonparametric version of the Hausman (1978) test. Lancaster (1990) and Van den Berg (2001) review other tests for misspecification of the mixing distribution, but not surprisingly, this is the first to use Theorem 2. We summarize this test in the following proposition and we apply the test in section 5. As before, let $\hat{\kappa}$ denote the estimator of this paper and let $\hat{\omega}$ denote an alternative estimator of the duration model with hazard $\theta(t|x,v) = v\exp(x(\lfloor t \rfloor)\beta_0)\lambda(t)$.

Proposition 1

Let the conditions of theorem 2 hold. Let $\sqrt{N}(\hat{\omega}-\kappa_0)$ converge to a normal distribution with mean zero and variance V_ω. Let V_ω be the asymptotic Cramer-Rao bound and let $\hat{V}_\omega - V_\omega = o_p(1)$, for some estimator $\hat{V}_\omega$. Let $\hat{V}_\kappa$ be the estimator of the asymptotic variance of $\hat{\kappa}$ calculated using the regular bootstrap. Then the limiting distribution of $\sqrt{N}(\hat{\omega}-\kappa_0)$ and $\sqrt{N}(\hat{\omega}-\hat{\kappa})$ has zero covariance and $\Upsilon = N\cdot(\hat{\omega}-\hat{\kappa})(\hat{V}_\kappa - \hat{V}_\omega)^{-1}T(\hat{\omega}-\hat{\kappa})$ has a chi-squared distribution with the number of degrees of freedom equal to the dimension of κ.

Proof: See appendix.

We apply this test in section 5 and reject the Hausman-Han-Meyer model that assumes a gamma distribution for the unobserved heterogeneity.

4. Gamma Mixing Distribution

Han and Hausman (1990) and Meyer (1990) use a flexible baseline hazard and model the unobserved heterogeneity as a gamma distribution[16]. Lancaster (1990) is very optimistic

[16] Ham and Rea (1987) also use a flexible baseline hazard but use a different mixing distribution.

that flexibility of the baseline hazard can somehow compensate for the restrictions of a gamma mixing distribution. In this section, we discuss the sensitivity of the estimators of the MPH model to misspecification of the mixing distribution. In particular, misspecifying the heterogeneity yields inconsistent estimators and having a flexible integrated baseline hazard $\Lambda(t)$ does not compensate for a failure to control for heterogeneity. We illustrate this using two examples.

Example 1:

Suppose we estimate the following hazard model: $\theta(t|v, X) = \phi^X \lambda(t)$. The function $\lambda(t)$ is nonparametric, and one could (incorrectly) conjecture that the flexibility of this function 'compensates' for the lack of unobserved heterogeneity. This model implies the following survivor function: $P(T \geq t|X) = \bar{F}(t \mid X) = \exp(-\phi^X \Lambda(t))$. Suppose we observe $\bar{F}(t \mid X)$ for $X = 0, 1$ and all $t \geq 0$. We define $\bar{F}_0(t) = \bar{F}(t \mid X = 0)$ and estimate

$$\hat{\Lambda}(t) = -\ln \bar{F}(t \mid X = 0).$$

For a given $\hat{\Lambda}(t) = -\ln \bar{F}(t \mid X = 0)$, the quasi maximum likelihood estimator of ϕ can be derived (see appendix), and it can be shown that

$$\operatorname*{plim}_{N\to\infty} \hat{\phi} = \frac{-1}{E[\ln\{\bar{F}_0(T)\}|X = 1]},$$

where $\bar{F}_0$ is the survival function for $X = 0$. Let the data be generated by the following model $\theta(t|v, X) = v\phi^X$ where $v \sim \text{Gamma}(\alpha, \alpha)$. Thus, $\bar{F}_0(t) = \left(1 + \frac{\Lambda(t)}{\alpha}\right)^{-\alpha}$ and $-\ln F_0(t) = \alpha \ln\left(1 + \frac{\Lambda(t)}{\alpha}\right)$. Note that $\phi^X \Lambda(T) = \frac{Z}{v}$, where Z has an exponential distribution with mean one. This yields

$$\operatorname*{plim}_{N\to\infty} \hat{\phi} = \frac{1}{E[\alpha \ln\{1 + Z/(\phi v \alpha\}]},$$

where $v \sim \text{Gamma}(\alpha, \alpha)$. Note that ϕ only appears in the denominator of the argument of a logarithmic function. This does not bode well for consistency. Using $N = 10,000$ we find the following:

True ϕ	True α	plim $\hat{\phi}$
$\phi = 2$	$\alpha = 1$	$\hat{\phi} = 1.46$
$\phi = 2$	$\alpha = 2$	$\hat{\phi} = 1.09$
$\phi = 10$	$\alpha = 1$	$\hat{\phi} = 4.04$
$\phi = 10$	$\alpha = 2$	$\hat{\phi} = 3.20$

.

Thus, the estimator for the regressor coefficient is inconsistent, despite the nonparametric baseline hazard.■

Example 2:

Suppose we estimate the following hazard model: $\theta(t|v,X) = ve^{X\beta_0}\lambda(t)$, where v has a gamma distribution. The function $\lambda(t)$ is nonparametric, and this time one could (incorrectly) conjecture that the flexibility of this function 'compensates' for the restrictive assumption that v has a gamma distribution. Suppose the data is generated by $\theta(t|v,X) = ve^{X}\lambda(t)$. Let $p(v)$ denote the density of v and let $p(v) = e^{c-v}$, $v \geq c$ and $c \geq 0$. Thus, v is an exponential random variable to which the nonnegative number c is added, and the true value of β equals one. Consider estimating this model under the assumption of gamma heterogeneity. Without loss of generality, we can write the integrated baseline hazard as follows:

$$\Lambda(t) = H(t)^d,$$

where $H(t)$ is unrestricted and $d > 0$. Horowitz (1996) and Chen (2002) show how to estimate $H(t)$ at the rate $\sqrt{N}$. Suppose that the conditions of Horowitz (1996) or Chen (2002) are satisfied and that one first estimates $H(t)$ using one of these methods. Estimating d is then like estimating a Weibull model. In the appendix, we show that the inconsistency of β does not depend on the distribution of the regressors. Using $N = 10,000$, we find the following:

c	β	γ_v	δ_v	$\beta; \gamma_v = 2, \delta_v = 1$
0	1	1	1	1
0.1	1.11	1.12	0.96	1.06
0.2	1.15	1.23	0.89	1.09
0.3	1.16	1.30	0.84	1.12
0.5	1.17	1.42	0.76	1.14
1	1.21	1.75	0.54	1.21
2	1.30	1.87	0.33	1.27

For $c = 0$, which is the correct specification, all parameters can be consistently estimated; the last column gives estimation results for β when $\gamma_v = 2$ and $\delta_v = 1$. The simulation results show that the inconsistencies increase with c.

■

Note that the asymptotic bias in the examples above does not depend on the shape of the hazard. The following lemma gives a reason for the asymptotic bias.

Lemma 1: *Let* $\theta(t \mid v, x) = ve^{x\beta_0}\lambda(t)$, *where* $v \perp x$. *Let* $v - c \mid T \geq 0 \sim Gamma(\gamma_v, \delta_v)$. *If* $c = 0$, *then* $\bar{F}(t|x)$ *decreases at a polynomial rate. If* $c > 0$, *then* $\bar{F}(t|x)$ *decreases at an exponential rate.*

The lemma states that the survivor probability as a function of time decreases at a polynomial rate if the unobserved heterogeneity distribution is a gamma distribution, and that the survivor probability decreases at an exponential rate if the unobserved heterogeneity distribution is a shifted gamma distribution. As the examples show, misspecification of the heterogeneity distribution cannot, in general, be corrected by a flexible baseline hazard. The estimator presented in this paper does not rely on specifying or estimating the heterogeneity distribution, which explains its better performance in terms of asymptotic bias and consistency.

5. Empirical Results

We estimate our new duration model on a sample of 15,491 males who received unemployment benefits beginning in 1998 in a data set called the Study of Unemployment Insurance Exhaustees public use data. The study was designed to examine the characteristics, labor market experiences, unemployment insurance (UI) program experiences, and reemployment service receipt of UI recipients.[17]

The study sample consists of UI recipients in 25 states who began their benefit year in 1998 and received at least one UI payment. It is designed to be nationally representative of UI exhaustees and non-exhaustees. The data description is:

> "The data come from the UI administrative records of the 25 sample states and telephone interviews conducted with a subsample of these UI recipients. Telephone interviews were conducted in English and Spanish between July 2000 and February 2001 using a two-stage process. For the first 16 weeks, all 25 participating states used mail, phone, and database methods to locate

[17] The description follows from http://www.upjohninst.org/erdc/uie/datasumm.html, which has further details about the sample design and results.

> sample members, who were then asked to complete the survey. The second stage, conducted in 10 of the sample states, added field staff to help locate non-responding sample members. The administrative data include the individual's age, race, sex, weekly benefit amount, first and last payment date, the state where benefits were collected, and whether benefits were exhausted." (op.cit.)

The survey data contain individual-level information about labor market and other activities from the time the person entered the UI system through the time of the interview. However, we limit our econometric study to the first 25 weeks of unemployment due to the recognized change in behavior in week 26 when UI benefits cease for a significant part of the sample (see, e.g., Han-Hausman (1990)). The data include information about the individual's pre-UI job, other income or assistance received, and demographic information.

We use two indicator variables, race and age over 50, in our index specification. We also use the replacement rate, which is the weekly benefit amount divided by the UI recipient's base period earnings. Lastly, we use the state unemployment rate of the state from which the individual received UI benefits during the period in which the individual filed for benefits. This variable changes over time. Table 1 gives the means and standard deviations for the variables we use in our empirical specification.

Table 1 here

We first estimate the unknown parameters of the model using the gamma heterogeneity specification of Han-Hausman (1990) and Meyer (1990) (HHM). This specification allows for a piecewise constant baseline hazard, which does not restrict the specification because unemployment duration is recorded on a weekly basis. However, it does impose a gamma heterogeneity distribution on the specification, which can lead to inconsistent estimates as we discussed above. We estimate the model using a gradient method and report the HHM estimates and bootstrap standard errors in Table 2. We calculate all the standard errors using the regular bootstrap and $10,000$ replications.

Table 2 here

The estimates of the parameters, as reported in Table 2, should not depend on how many weeks of data we use (6, 13 or 24 weeks). However, the coefficients differ significantly. We find significant evidence of heterogeneity in the two larger samples, but in the 6 period sample, we do not estimate significant heterogeneity. We also find the expected negative estimates for all of the coefficients, with the state unemployment rate a significant factor in affecting the probability of exiting unemployment. When comparing the estimates of the parameters across the 3 samples, the scaling changes depending on the variance of the estimated gamma distribution. Thus, the ratios of the coefficients should be compared. The ratios of the coefficients across samples remain similar, with the results for the 13 period and 24 period samples very close to each other.

We now turn to an estimate of the new duration specification, which does not require estimation of a heterogeneity distribution, using the same samples as above. Optimization of the objective function can now create a problem because of its lack of smoothness. Usual Newton-type gradient methods or conjugate gradient (simplex) methods do not work in this situation. To date, we have found that generalized pattern search algorithms perform best.[18] We use the pattern search routine from Matlab to estimate the parameters; see Appendix 7 for further details about our computational approach. The basic idea is to begin with the gamma heterogeneity estimates and to construct a "bounding box" of 3 standard deviations around each parameter estimate. We then find new estimates and increase the bounding box until we do not find an increase in the objective function. The routine converges relatively rapidly. We calculate all the standard errors using the regular bootstrap and $10,000$ replications. In Table 3, we give the estimates from the new duration model. We also check our pattern search results using a genetic optimization approach that is also discussed in the appendix. The genetic optimization approach has

[18] Further research would be helpful here. We have also used gradient algorithms on a smoothed objective function to obtain initial estimates and then employed Nelder-Mead routines to find the optimum. However, the pattern search algorithms appear to work best. See Audet and Dennis (2003) for a recent survey of pattern search algorithms.

the advantage of not depending on initial values. However, it has the disadvantage of taking much longer to solve, so it cannot be used feasibly to bootstrap the results to estimate the standard errors. However, the results of the pattern search algorithm and the genetic optimization algorithm are very similar as we describe in the appendix.

Table 3 here

Again we find that all of the estimated coefficients have the expected negative signs. The coefficients are also estimated with a high degree of statistical precision, although this finding may be a result of our large sample size of 15,491 individuals. We again find that the ratios of coefficients remain relatively stable across the three different samples with the exception of the replacement rate, which becomes increasingly larger with respect to the state unemployment rate as the sample length increases. The change in the estimated coefficient for the replacement rate for the 24 week sample appears to arise because most recipients' unemployment insurance terminates after 26 weeks. Han-Hausman (1990) find a significant change in behavior at week 26. As individuals start to approach week 26 the size of the replacement rate has a diminished effect on their behavior as they foresee the end of their unemployment benefits beginning to draw near.

In Figures 1 and 2, we plot the survival curves for the 13 week and 24 week gamma heterogeneity estimates and for the estimates from the new model. These figures also have a 95% confidence band (point by point) for the survival curves. We fit the survival curves using a second order local polynomial estimator which takes account of the standard deviations of the estimated period coefficients in Table 2 and 3.[19] The estimated local polynomial survival curves fit the data well for all specifications.

Figure 1 here

[19] We explain our approach in more detail in the appendix.

Figure 2 here

We find that the new model gives extremely similar results for the 6 period data and the 13 period data. Indeed, a Hausman (1978) specification test on the slope coefficients is 0.42 with 4 degrees of freedom. Thus, we find that the new model is not sensitive to the number of periods used to estimate the model. For the 24 period model, we find the coefficients again very close to the other results except for the coefficient of the replacement rate. A Hausman test now rejects the equality of the slope coefficients with a value of 234.3, based essentially on the change in the replacement rate coefficient (comparing 24 to 13 weeks). However, since most individuals' unemployment benefits run out in the 26th week, the change in the estimated coefficient is likely because of unmodeled dynamics at the point of benefit exhaustion. Lastly, if we test the ratios of the gamma heterogeneity model versus the new duration model, we do not reject that the ratios are the same for 6 periods with a test value of 3.5; we marginally reject equality of the coefficient ratios for 13 periods with a test value of 6.2; and we do reject equality of the coefficient ratios for 24 weeks with a test value of 12.4. Thus, the new duration model does find differences from the previous gamma heterogeneity model. The new duration model also has the advantage that the absolute values of the estimated coefficients are not sensitive to the length of the data period, while the gamma heterogeneity model does not have this property.

The main difference we find between the results of the gamma heterogeneity survival curves and those of the semi-parametric survival curves is that the gamma heterogeneity survival curves are initially steeper. Thus, the gamma heterogeneity results predict a higher probability of exiting unemployment in the early periods than do the semi-parametric results. However, again the differences are not substantial. We reject equality of the survival curves due to the extremely small standard errors we estimate with our very large sample.

6. Conclusion

Since Lancaster (1979), it has been recognized that it is important to account for unobserved heterogeneity in models for duration data. Failure to account for unobserved

heterogeneity makes the estimated hazard rate decrease more with the duration than the hazard rate of a randomly selected member of the population. In this paper, we derive a new estimator for the MPH model that allows for a nonparametric baseline hazard and time-varying regressors. By using time-varying regressors, we are able to estimate the regression coefficients, instead of estimates only up to scale as in some of the previous literature. We also do not require explicit estimation of the heterogeneity distribution in estimating the baseline hazard and regression coefficients. The baseline hazard rate is nonparametric and the estimator of the integrated baseline hazard rate converges at the regular rate, $\sqrt{N}$, where N is the sample size. This is the same rate as for a duration model without heterogeneity. The regressor parameters also converge at the regular rate. Also, a hazard model is a natural framework for time-varying regressors. In particular, if a flow or a transition probability depends on a regressor that changes with time, a hazard model avoids the curse of dimensionality that would arise from interacting the regressors at each point in time with one another. A nice feature of the new estimator is that it allows the durations to be measured on a finite set of points. Such discrete measurement of durations is important in economics; for example, unemployment is often measured in weeks. We also propose a new test to detect unobserved heterogeneity and, also, misspecified unobserved heterogeneity. The test is a nonparametric version of the Hausman (1978) test. We use it in the application to test the gamma distribution assumption in the Han-Hausman-Meyer model and reject the assumption that the unobserved heterogeneity has a gamma distribution.

Appendix 1: Derivation of the Objective Function

In order to simplify notation, we write $Z_i(l)$ and $Z_j(k)$ instead of $Z_i(l;\kappa)$ and $Z_j(k;\kappa)$.

$$\begin{aligned}
Q(\kappa) &= \frac{1}{N(N-1)}\sum_i\sum_j\sum_{l=1}^K\sum_{k=1}^K[1\{T_i\geq l\}-1\{T_j\geq k\}]1\{Z_i(l)<Z_j(k)\}\\
&= \frac{1}{N(N-1)}\sum_i\sum_j\sum_{l=1}^K[1\{T_i\geq l\}\sum_{k=1}^K 1\{Z_i(l)<Z_j(k)\}\\
&- \frac{1}{N(N-1)}\sum_i\sum_j\sum_{l=1}^K 1\{T_j\geq k\}\sum_{k=1}^K 1\{Z_i(l)<Z_j(k)\}\\
&= \frac{\sum_i}{N}\sum_{l=1}^K 1\{T_i\geq l\}\frac{\sum_j}{N-1}\sum_{k=1}^K[1\{Z_i(l)<Z_j(k)\}-1\{Z_i(l)>Z_j(k)\}]\\
&= \frac{\sum_i}{N}\sum_{l=1}^K 1\{T_i\geq l\}\frac{\sum_j}{N-1}\sum_{k=1}^K[1-2*1\{Z_j(k)<Z_i(l)\}]
\end{aligned}$$

with probability one. Let

$$\hat{F}_Z\{Z_i(l)\}=\frac{\sum_j}{N-1}\frac{\sum_{k=1}^K}{K}1\{Z_j(k)<Z_i(l)\},$$

i.e. $\hat{F}_Z\{(.)$ is an estimator of the cumulative distribution function $F_Z(.)$, i.e.$\hat{F}_Z\{Z_i(l)\}=F_Z\{Z_i(l)\}+o_p(1)$, for $l=1,...,K$ and $i=1,...,N$. Using $\hat{F}_Z(.)$ yields

$$Q(\kappa)=\frac{\sum_i}{N}\sum_{l=1}^K 1\{T_i\geq l\}K[1-2\hat{F}_Z\{Z_i(l)\}].$$

Appendix 2: Proof of Theorem 1

The first lemma shows that one can estimate β up to scale by using the data of period one (i.e. $K=1$). As before, let β_0 denote the true value of the parameter and $\beta_{0,1}$ be the first element of the vector β_0 (we only use the subscript 'zero' when there is a risk of confusion between an element of the parameter space and the true value). Note the normalization in the text, $\delta_1=0$, so that the objective function is only a function of β if $K=1$. In particular, let

$$Q(\beta,\delta_1=0)=\frac{1}{N(N-1)}\sum_i\sum_j[1\{T_i\geq 1\}-1\{T_j\geq 1\}]1\{Z_i(1;\beta,\delta_1=0)<Z_j(1;\beta,\delta_1=0)\}.$$

Lemma A1: *Let assumption 1(i)–(v) and (vii) hold. Let $\beta_0\in\Theta_\beta$, which is compact, $K=1$, and $\hat{\beta}/|\hat{\beta}_1|=\underset{\beta\in\Theta_\beta}{\operatorname{argmin}}\ Q(\beta)$. Then*

$$\hat{\beta}/|\hat{\beta}_1|\underset{p}{\to}\beta_0/|\beta_{0,1}|$$

Proof: The same reasoning as in the main text implies that the true values $\{\beta_0\}$ yield *a* minimum of the expectation of the objective function. We now show that $\beta_0/|\beta_{0,1}|$ yields a *unique* minimum, i.e. $\beta_0/|\beta_{0,1}| = \underset{\beta \in \Theta_\beta}{argmin}\ E\{Q(\beta)\}$. Let W denote the regressors in period 1. Note that the support of W is not contained in any proper linear subspace of $\mathbb{R}^M$. This implies that $E\{WW'\}$ is positive definite (e.g. see Newey and McFadden (1994, page 2125)). Therefore, for any $\gamma^* \neq \gamma$, $W'(\gamma - \gamma^*) \neq 0$ on a set with positive probability. One needs that $1\{W'\gamma < 0\} \neq 1\{W'\gamma^* < 0\}$ on a set with positive probability. To see that this is the case, note that the first component of W, conditional on the other regressors, is continuously distributed with an infinite support by assumption. Using $\gamma = \beta/|\beta_1|$ gives that $E\{Q(\beta)\}$ is minimized at $\beta_0/|\beta_{0,1}|$. Thus, β is identified up to scale. Also note that the conditions of Newey and McFadden (1994, theorem 2.1 and lemma 2.8) are satisfied and so that $\hat{\beta}/|\hat{\beta}_1| \underset{p}{\to} \beta_0/|\beta_{0,1}|$. Q.E.D.

The next lemma shows that one can identify the scale parameter, β_1, and δ_2 by using data from period 1 and 2. For simplicity, the next lemma assumes that x_{it} is a scalar for all i and all t.

Lemma A2: *Let assumptions 1-2 hold. Let $K = 2$ and let x_{it} be a scalar for all i and all t. Then*

$$\{\hat{\beta}, \hat{\delta}\} \underset{p}{\to} \{\beta_0, \delta_0\}$$

Proof: We first establish identification and then show that the estimator converges in probability. Note that assumptions 1-2 hold, so that a regressor stays constant over time with positive probability. Note that this model only has two scalar parameters, β and δ_2. Without loss of generality, let $\beta_0 > 0$ (if $\beta_0 < 0$, multiply x by -1). Consider the following reparametrization[20]: $\delta_2 = \ln(e^{\beta c} - 1)$ for some $c > 0$. The same reasoning as in the main text implies that the true values $\{\beta_0, c_0\}$ yield *a* minimum of the expectation of the objective function. We now show that $\{\beta_0, c_0\}$ yields a *unique* minimum, i.e. $\{\beta_0, c_0\} = \underset{\beta, c \in \Theta}{argmin}\ E\{Q(\beta, c)\}$. In particular, we first show that c_0 yields a unique minimum for those individuals whose regressors do not change from period one to period two

[20] The reparametrization simplifies the identification proof; see Woutersen (2002) for an overview of such techniques.

so that $E\{Q(\beta, c_0)\} < E\{Q(\beta, c)\}$ for any $\beta, c \in \Theta$ and $c \neq c_0$. Thus, the advantage of the reparametrization is that we can identify c without having to identify β. We then show, using the assumption that some regressors vary with time, that $E\{Q(\beta_0, c_0)\} < E\{Q(\beta, c)\}$ for any $\beta, c \in \Theta$ and $c \neq c_0, \beta \neq \beta_0$. Consider the expectation of the contribution of the pair $i \neq j$ to the objective function after conditioning on $x_{i1} = x_{i2}$,

$$\begin{aligned} & E[\{e^{-vZ_i(l=2;\beta_0,\delta_{0,2})} - e^{-vZ_j(k=1;\beta_0,\delta_{0,2})}\} \cdot 1\{Z_i(l=2;\beta,\delta) < Z_j(k=1;\beta,\delta)\}|x_i, x_j, x_{i1} = x_{i2}] \\ = \; & E([\exp\{-v(e^{x_{i1}\beta_0} + e^{x_{i1}\beta_0+\delta_{0,2}})\} - \exp\{-ve^{x_{j1}\beta_0}\}] \cdot 1\{e^{x_{i1}\beta} + e^{x_{i1}\beta+\delta} < e^{x_{j1}\beta}\}]|x_i, x_j, x_{i1} = x_{i2}). \end{aligned}$$

Using $\delta_2 = \ln(e^{c\beta} - 1)$ for some $c > 0$ yields $e^{x_{i1}\beta} + e^{x_{i1}\beta+\delta} = e^{x_{i1}\beta+c\beta}$. Thus,

$$\begin{aligned} & E[\{e^{-vZ_i(l=2;\beta_0,\delta_{0,2})} - e^{-vZ_j(k=1;\beta_0,\delta_{0,2})}\} \cdot 1\{Z_i(l=2;\beta,\delta) < Z_j(k=1;\beta,\delta)\}|x_i, x_j, x_{i1} = x_{i2}] \\ = \; & E([\exp\{-v(e^{x_{i1}\beta_0+c_0\beta_0})\} - \exp\{-ve^{x_{j1}\beta_0}\}] \cdot 1\{e^{x_{i1}\beta+c\beta} < e^{x_{j1}\beta}\}]x_i, x_j, x_{i1} = x_{i2}) \\ = \; & E([\exp\{-v(e^{x_{i1}\beta_0+c_0\beta_0})\} - \exp\{-ve^{x_{j1}\beta_0}\}] \cdot 1\{c - (x_{j1} - x_{i1}) < 0\}]|x_i, x_j, x_{i1} = x_{i2}) \qquad (6) \\ = \; & E([\exp\{-v(e^{x_{i1}\beta_0+c_0\beta_0})\} - \exp\{-ve^{x_{j1}\beta_0}\}] \cdot 1\{c - x_{ij} < 0\}]|x_i, x_j, x_{i1} = x_{i2}), \end{aligned}$$

where $x_{ij} = x_{j1} - x_{i1}$. Note that $E[\exp\{-v(e^{x_{i1}\beta_0+c_0\beta_0})\} - \exp\{-ve^{x_{j1}\beta_0}\}] < 0$ if and only if $c - x_{ij} < 0$ and that $E[\exp\{-v(e^{x_{i1}\beta_0+c_0\beta_0})\} - \exp\{-ve^{x_{j1}\beta_0}\}] > 0$ if and only if $c - x_{ij} > 0$. Also, assumption 2 implies that $\{c_0 - x_{ij}\}$ has support around zero, so c_0 is identified, i.e. $E\{Q(\beta, c_0)\} < E\{Q(\beta, c)\}$ for any $\beta, c \in \Theta$ and $c \neq c_0$. Using this result, we now show that $E\{Q(\beta_0, c_0)\} < E\{Q(\beta, c)\}$ for any $\beta, c \in \Theta$ and $c \neq c_0, \beta \neq \beta_0$. Define

$$\begin{aligned} H_{ij}(\beta, c) &= e^{x_{i1}\beta} + e^{x_{i2}\beta+\delta_2} - e^{x_{j1}\beta} \qquad (7) \\ &= e^{x_{i1}\beta} + e^{x_{i2}\beta+c\beta} - e^{x_{i2}\beta} - e^{x_{j1}\beta} \end{aligned}$$

using $\delta_2 = \ln(e^{c\beta} - 1)$. Dividing by $e^{x_{i1}\beta}$ yields

$$H^*_{ij}(\beta, c) = 1 + e^{(x_{i2}-x_{i1}+c)\beta} - e^{(x_{i2}-x_{i1})\beta} - e^{(x_{j1}-x_{i1})\beta}.$$

Differentiating with respect to β gives

$$\frac{\partial H^*_{ij}(\beta, c)}{\partial \beta} = (x_{i2} - x_{i1} + c)e^{(x_{i2}-x_{i1}+c)\beta} - (x_{i2} - x_{i1})e^{(x_{i2}-x_{i1})\beta} - (x_{j1} - x_{i1})e^{(x_{j1}-x_{i1})\beta}.$$

Let $P(T_i \geq 2|x_i) > P(T_j \geq 1|x_j)$ so that $E[\exp\{-v(e^{x_{i1}\beta_0} + e^{x_{i2}\beta_0+\beta_0 c_0} - e^{x_{i2}\beta_0})\}|x_i] > E[\exp\{-v(e^{x_{j1}\beta_0})\}|x_j]$. This implies that $H_{ij}(\beta_0, c_0) = e^{x_{i1}\beta_0} + e^{x_{i2}\beta_0+\beta_0 c_0} - e^{x_{i2}\beta_0} -$

$e^{x_{j1}\beta_0} \leq 0$ and that $H^*_{ij}(\beta_0, c_0) = 1 + e^{(x_{i2}-x_{i1}+c_0)\beta_0} - e^{(x_{i2}-x_{i1})\beta_0} - e^{(x_{j1}-x_{i1})\beta_0} < 0$. Suppose that $x_{i2} - x_{i1} < 0$ so that $1 - e^{(x_{i2}-x_{i1})\beta_0} > 0$ for any value of $\beta_0 > 0$. This implies, using $H^*_{ij}(\beta_0, c_0) < 0$, that $e^{(x_{i2}-x_{i1}+c_0)\beta_0} < e^{(x_{j1}-x_{i1})\beta_0}$ so that $(x_{i2} - x_{i1} + c_0) < (x_{j1} - x_{i1})$. This implies that $\frac{\partial H^*_{ij}(\beta,c_0)}{\partial \beta} < 0$ for all $\beta > 0$ so that $H^*_{ij}(\beta, c_0) < H^*_{ij}(\beta_0, c_0)$ if $\beta > \beta_0$ and $H^*_{ij}(\beta, c_0) > H^*_{ij}(\beta_0, c_0)$ if $\beta < \beta_0$. In particular, given assumption 1-2, for those values of the regressors for which $P(T_i \geq 2|x_i, x_{i1} > x_{i2}) > P(T_j \geq 1|x_j)$ and $x_{i2} - x_{i1} < 0$, the conditional expectations of the contributions to the objective functions,

$$\{P(T_i \geq 2|x_i, x_{i1} > x_{i2}) - P(T_j \geq 1|x_j)\} * 1\{H^*_{ij}(\beta, c_0) < 0\},$$

are minimized for any value of β for which $\beta \geq \beta_0$.

Now suppose $P(T_i \geq 2|x_i, x_{i1} > x_{i2}) < P(T_j \geq 1|x_j)$. In this case, $E[\exp\{-v(e^{x_{i1}\beta_0} + e^{x_{i2}\beta_0+\beta_0 c_0} - e^{x_{i1}\beta_0})\}] < E[\exp\{-v(e^{x_{j1}\beta_0})\}]$. This implies that $H_{ij}(\beta_0, c_0) = e^{x_{i1}\beta_0} + e^{x_{i2}\beta_0+\beta_0 c_0} - e^{x_{i2}\beta_0} - e^{x_{j1}\beta_0} > 0$ and that $H^{**}(\beta_0, c_0) = e^{(x_{i1}-x_{i2})\beta_0} + e^{\beta_0 c_0} - 1 - e^{(x_{j1}-x_{i2})} > 0$. Again, suppose that $x_{i2} - x_{i1} < 0$ so that $e^{(x_{i1}-x_{i2})\beta_0} - 1 > 0$ for any value of $\beta_0 > 0$. This implies that $e^{c_0\beta_0} > e^{(x_{j1}-x_{i2})\beta_0}$ so that $c_0 > (x_{j1} - x_{i2})$. This implies that $\frac{\partial H^{**}(\beta,c_0)}{\partial \beta} > 0$. Similar reasoning as above implies that the conditional expectations of the contributions to the objective functions,

$$\{P(T_i \geq 2|x_i, x_{i1} > x_{i2}) - P(T_j \geq 1|x_j)\} * 1\{H^*(\beta, c_0) < 0\},$$

are minimized for any value of β for which $\beta \leq \beta_0$. Thus, β_0 is identified if $x_{i2} - x_{i1} < 0$. A similar reasoning applies if $x_{i2} - x_{i1} > 0$, so β_0 is identified under the assumptions. Given that $\delta_2 = \ln(e^{c\beta} - 1)$, identification of $\{\beta, c\}$ is equivalent to identification of $\{\beta, \delta\}$. Also note that the conditions of Newey and McFadden (1994, theorem 2.1 and lemma 2.8) are satisfied and so that $\{\hat{\beta}, \hat{\delta}\} \underset{p}{\rightarrow} \{\beta_0, \delta_0\}$. Q.E.D.

Proof of **Theorem 1**:

The same reasoning as in the main text implies that the true values $\{\beta_0, \delta_0\}$ yield *a* minimum of the expectation of the objective function. We now show that $\{\beta_0, \delta_0\}$ yields a *unique* minimum, i.e. $\{\beta_0, \delta_0\} = \underset{\beta,\delta\in\Theta}{argmin}\ E\{Q(\beta, \delta)\}$ using lemma 1 and lemma 2. We first consider $K = 3$. Lemma A1 and lemma A2 imply that $\{\beta, \delta_2\}$ are identified so that

δ_3 is the only remaining parameter to be identified. Consider the objective function,

$$\begin{aligned} Q(\kappa) &= \frac{1}{N(N-1)} \sum_i \sum_j \sum_{l=1}^{K} \sum_{k=1}^{K} Q_{ijlk} \\ &= \frac{1}{N(N-1)} \sum_i \sum_j \sum_{l=1}^{K} \sum_{k=1}^{K} [1\{T_i \geq l\} - 1\{T_j \geq k\}] 1\{Z_i(l;\kappa) < Z_j(k;\kappa)\} \end{aligned}$$

and consider summing all contributions, Q_{ijlk} of the objective function for which $k = K = 3$. That is,

$$\frac{1}{N(N-1)} \sum_i \sum_j \sum_{l=1}^{K} \sum_{k=3}^{K=3} [1\{T_i \geq l\} - 1\{T_j \geq k\}] 1\{Z_i(l;\kappa) < Z_j(k;\kappa)\}.$$

The same reasoning as in lemma A1 (including using the full support condition for the first regressor in the first period) implies that δ_3 is identified. We can use a similar argument for $K = 4, 5$ etc. so that $\{\beta_0, \delta_0\}$ is the unique minimum of $E\{Q(\beta, \delta)\}$.

Next, we show convergence in probability. Define

$$\begin{aligned} Q_0(\beta, \delta) &= E\{Q(\beta, \delta)\} \\ &= E[E\{Q(\beta, \delta)|Z\}] \\ &= E[\frac{\sum_i}{N} \sum_{l=1}^{K} E_v\{e^{-vZ_i(l;\kappa)}|Z_i(l;\kappa)\} \sum_{k=1}^{K} [2 * F_Z(Z_i(l;\kappa)) - 1]], \end{aligned}$$

where F_Z is the cdf of $Z_i(l;\kappa)$ for $l = 1, ..., K$ and $i = 1, ..., N$. The function $Q_0(\beta, \delta)$ is continuous and minimized at the true value of the parameters. The function $Q(\beta, \delta)$ is stochastically equicontinuous, and the conditions of Newey and McFadden (1994, lemma 2.8) are satisfied, so that $Q(\beta, \delta)$ converges uniformly to $EQ(\beta, \delta)$. Moreover, Θ is assumed to be compact and the data are i.i.d., so consistency follows from Newey and McFadden (1994, theorem 2.1). Note that these arguments do not require that there should be unobserved heterogeneity; they still hold if all individuals have the same value of v. Q.E.D.

Appendix 3: Proof of Theorem 2: Asymptotic Normality

In order to simplify notation we write $Z_i(l)$ and $Z_j(k)$ instead of $Z_i(l;\kappa)$ and $Z_j(k;\kappa)$. We prove the theorem by applying theorem 1 and theorem 2 by Sherman (1993).[21] Let $||\cdot||$ denote the L_2 norm. Sherman's (1993) condition (i) in theorem 1 requires that there exists a neighborhood $\mathcal{N}$ of κ_0 and a constant $C > 0$ for which

$$E\{Q(\kappa)\} - E\{Q(\kappa_0)\} \leq C||\kappa - \kappa_0||^2$$

for all κ in $\mathcal{N}$. Note that, in our case, the expectation of the objective function is twice continuously differentiable and that κ_0 uniquely minimizes $E\{Q(\kappa)\}$. Therefore we can use the following Taylor approximation around the minimum. In particular, consider an arbitrary $s \in \mathbb{R}^{\dim(\kappa)}$ and consider the second order Taylor expansion,

$$E\{Q(\kappa_0 + \eta s)\} = E\{Q(\kappa_0)\} + s'\nabla_{\kappa\kappa}Q_0(\kappa_0)s\eta^2 + o(\eta^2).$$

The Hessian $\nabla_{\kappa\kappa}Q_0(\kappa_0)$ is negative definite so that it is possible to choose $\mathcal{N}$ such that $s'\nabla_{\kappa\kappa}Q_0(\kappa_0)s$ dominates the higher order terms. Thus, Sherman (1993, theorem 1, condition (i)) is satisfied. Condition (ii) in theorem 1 requires that

$$Q(\kappa) = Q(\kappa_0) + E\{Q(\kappa)\} - E\{Q(\kappa_0)\} + O_p(||\kappa - \kappa_0||/\sqrt{N}) + o_p(||\kappa - \kappa_0||^2) + O_p(1/N)$$

uniformly over $o_p(1)$ neighborhoods of κ_0. In order to show that this assumption holds, define

$$\triangle_{1,N}(\kappa, \kappa_0) = Q(\kappa) - E\{Q(\kappa)|X\} - [Q(\kappa_0) - E\{Q(\kappa_0)|X\}], \text{ and}$$

$$\triangle_{2,N}(\kappa, \kappa_0) = E\{Q(\kappa)|X\} - E\{Q(\kappa)\} - [E\{Q(\kappa_0)|X\} - E\{Q(\kappa_0)\}],$$

where X denotes the data on the regressors for all individuals. Note that

$$Q(\kappa) - Q(\kappa_0) - E\{Q(\kappa)\} + E\{Q(\kappa_0)\} = \triangle_{1,N}(\kappa, \kappa_0) + \triangle_{2,N}(\kappa, \kappa_0)$$

so that we have to show that $\triangle_{1,N}(\kappa, \kappa_0) + \triangle_{2,N}(\kappa, \kappa_0)$ is $O_p(||\kappa - \kappa_0||/\sqrt{N}) +$ $o_p(||\kappa - \kappa_0||^2) + O_p(1/N)$ uniformly over $o_p(1)$ neighborhoods of κ_0. Consider $\triangle_{1,N}(\kappa, \kappa_0)$

[21] Note that Sherman (1993) normalizes $Q(\kappa_0)$, $E\{Q(\kappa_0)\}$. and κ_0 to be zero. We do not use these normalizations and, instead, use $Q(\kappa) - Q(\kappa_0)$, $E\{Q(\kappa)\} - E\{Q(\kappa_0)\}$ and $(\kappa - \kappa_0)$.

and note that

$$\begin{aligned}\triangle_{1,N}(\kappa,\kappa_0) &= \frac{\sum_i}{N}\sum_{l=1}^{K}[1(T_i\geq l)-E\{1(T_i\geq l)|X_i\}]K[-2\hat{F}_Z\{Z_i(l)\}+2\hat{F}_{Z_0}\{Z_{0,i}(l)\}] \\ &= \frac{\sum_i}{N}\sum_{l=1}^{K}[1(T_i\geq l)-E\{1(T_i\geq l)|X_i\}]K[-2F_Z\{Z_i(l)\}+2F_{Z_0}\{Z_{0,i}(l)\}]+ \\ &\quad +\frac{\sum_i}{N}\sum_{l=1}^{K}[1(T_i\geq l)-E\{1(T_i\geq l)|X_i\}]K[-2\hat{F}_Z\{Z_i(l)\} \\ &\quad +2F_Z\{Z_i(l)\}+2\hat{F}_{Z_0}\{Z_{0,i}(l)\}-2F_{Z_0}\{Z_{0,i}(l)\}].\end{aligned}$$

where X_i denotes the data on the regressors of individual i. Let $\triangle_{1,N}^{small}(\kappa,\kappa_0)$ denote the last term, i.e.

$$\begin{aligned}\triangle_{1,N}^{small}(\kappa,\kappa_0) &= \frac{\sum_i}{N}\sum_{l=1}^{K}[1(T_i\geq l)-E\{1(T_i\geq l)|X_i\}]K[-2\hat{F}_Z\{Z_i(l)\}+2F_Z\{Z_i(l)\} \\ &\quad +2\hat{F}_{Z_0}\{Z_{0,i}(l)\}-2F_{Z_0}\{Z_{0,i}(l)\}].\end{aligned}$$

Note that $E[1(T_i\geq l)-E\{1(T_i\geq l)|X_i\}|X]=0$ so that $E\{\triangle_{1,N}^{small}(\kappa,\kappa_0)\}=E[E\{\triangle_{1,N}^{small}(\kappa,\kappa_0)|X\}]=0$. Next, note that for $i\neq j$, covariance$\{1(T_i\geq l),1(T_j\geq k)|X\}=0$ since the data is i.i.d. By the usual formula for the variation[22], we have that

$$\begin{aligned}var\{\triangle_{1,N}^{small}(\kappa,\kappa_0)\} &= E[var\{\triangle_{1,N}^{small}(\kappa,\kappa_0)|X\}]+var[E\{\triangle_{1,N}^{small}(\kappa,\kappa_0)|X\}] \\ &= \frac{1}{N^2}E[\sum_i var\{\sum_{l=1}^{K}1(T_i\geq l)K[-2\hat{F}_Z\{Z_i(l)\}+2F_Z\{Z_i(l)\} \\ &\quad +2\hat{F}_{Z_0}\{Z_{0,i}(l)\}-2F_{Z_0}\{Z_{0,i}(l)\}]|X\}].\end{aligned}$$

Note that K is fixed and that var$[\hat{F}_Z\{Z_i(l)\}]$ and cov$[\hat{F}_Z\{Z_i(l)\},\hat{F}_{Z_0}\{Z_{0,i}(l)\}]$ are differentiable so that a first order Taylor expansion around κ_0 gives that $var\{\triangle_{1,N}^{small}(\kappa,\kappa_0)\}=O(\frac{||\kappa-\kappa_0||}{N^2})$ so that $\triangle_{1,N}^{small}(\kappa,\kappa_0)=O_p(\frac{||\kappa-\kappa_0||^{1/2}}{N})$. Also note that $var\{\triangle_{1,N}^{small}(\kappa,\kappa_0)\}$ is $O(\frac{1}{N^2})$ for any fixed value of κ and κ_0. Next, define

$$\begin{aligned}\triangle_{1,N}^{main}(\kappa,\kappa_0) &= \triangle_{1,N}(\kappa,\kappa_0)-\triangle_{1,N}^{small}(\kappa,\kappa_0) \\ &= \frac{\sum_i}{N}\sum_{l=1}^{K}[1(T_i\geq l)-E\{1(T_i\geq l)|X_i\}]K[-2F_Z\{Z_i(l)\}+2F_{Z_0}\{Z_{0,i}(l)\}].\end{aligned}$$

[22] $var(A)=E(var(A|B))+var(E(A|B)$ for any random variables A,B that have finite second moments.

Using $var\{\triangle_{1,N}^{main}(\kappa,\kappa_0)\} = E[var\{\triangle_{1,N}^{main}(\kappa,\kappa_0)|X\}] + var[E\{\triangle_{1,N}^{main}(\kappa,\kappa_0)|X\}]$ and a first order Taylor expansion around κ_0 gives

$$\triangle_{1,N}^{main}(\kappa,\kappa_0) = \frac{\sum_i}{N}\sum_{l=1}^{K}[1(T_i \geq l) - E\{1(T_i \geq l)|X_i\}]K[-2\frac{\partial F_Z\{Z_i(l)\}}{\partial \kappa}'(\kappa-\kappa_0)] + o_p(||\kappa-\kappa_0||/\sqrt{N}).$$

The first term of $\triangle_{1,N}^{main}(\kappa,\kappa_0)$ drives the normality result in theorem 2. Next, consider

$$\begin{aligned}
\triangle_{2,N}(\kappa,\kappa_0) &= E\{Q(\kappa)|X\} - E\{Q(\kappa)\} - [E\{Q(\kappa_0)|X\} + E\{Q(\kappa_0)\}] \\
&= \frac{\sum_i}{N}\sum_{l=1}^{K}(E\{1(T_i \geq l)|X_i\}K[-2\hat{F}_Z\{Z_i(l)\} + 2\hat{F}_{Z_0}\{Z_{0,i}(l)\}]) \\
&\quad - \frac{\sum_i}{N}E\{\sum_{l=1}^{K}(E\{1(T_i \geq l)|X_i\}K[-2\hat{F}_Z\{Z_i(l)\} + 2\hat{F}_{Z_0}\{Z_{0,i}(l)\}])\}.
\end{aligned} \tag{8}$$

We again use arguments based on conditional expectations. Therefore, it is useful to note that $\hat{F}_Z\{Z_i(l)\}$ only depends on $Z_i(l)$ through its argument. That is

$$\begin{aligned}
\hat{F}_Z\{Z_i(l)\} &= \frac{\sum_j}{N-1}\frac{\sum_{k=1}^{K}}{K}1\{Z_j(k) < Z_i(l)\} \\
&= \frac{\sum_{j\neq i}}{N-1}\frac{\sum_{k=1}^{K}}{K}1\{Z_j(k) < Z_i(l)\} + \frac{l-1}{(N-1)K}.
\end{aligned}$$

We can use this to simplify the last term in equation (8). In particular,

$$\begin{aligned}
&\frac{\sum_i}{N}E\{\sum_{l=1}^{K}(E\{1(T_i \geq l)|X_i\}K[-2\hat{F}_Z\{Z_i(l)\} + 2\hat{F}_{Z_0}\{Z_{0,i}(l)\}])\} \\
&\quad = \frac{\sum_i}{N}E[\sum_{l=1}^{K}E\{(E\{1(T_i \geq l)|X_i\}K[-2\hat{F}_Z\{Z_i(l)\} + 2\hat{F}_{Z_0}\{Z_{0,i}(l)\}])|X_i\}] \\
&\quad = \frac{\sum_i}{N}E\{\sum_{l=1}^{K}(E\{1(T_i \geq l)|X_i\}K[-2F_Z\{Z_i(l)\} + 2F_{Z_0}\{Z_{0,i}(l)\}])\}.
\end{aligned}$$

Let

$$\begin{aligned}
\triangle_{2,N}^{A}(\kappa,\kappa_0) &= \frac{\sum_i}{N}\sum_{l=1}^{K}(E\{1(T_i \geq l)|X_i\}K[-2\hat{F}_Z\{Z_i(l)\} + 2\hat{F}_{Z_0}\{Z_{0,i}(l)\}]) \\
&\quad - \frac{\sum_i}{N}\sum_{l=1}^{K}(E\{1(T_i \geq l)|X_i\}K[-2F_Z\{Z_i(l)\} + 2F_{Z_0}\{Z_{0,i}(l)\}]), \text{ and}
\end{aligned} \tag{9}$$

$$\begin{aligned}
\triangle_{2,N}^{B}(\kappa,\kappa_0) &= \frac{\sum_i}{N}\sum_{l=1}^{K}(E\{1(T_i \geq l)|X_i\}K[-2F_Z\{Z_i(l)\} + 2F_{Z_0}\{Z_{0,i}(l)\}]) \\
&\quad - \frac{\sum_i}{N}E\{\sum_{l=1}^{K}(E\{1(T_i \geq l)|X_i\}K[-2F_Z\{Z_i(l)\} + 2F_{Z_0}\{Z_{0,i}(l)\}])\}.
\end{aligned} \tag{10}$$

and note that $\triangle_{2,N}(\kappa,\kappa_0) = \triangle^A_{2,N}(\kappa,\kappa_0) + \triangle^B_{2,N}(\kappa,\kappa_0)$. Thus,

$$\begin{aligned}\triangle^A_{2,N}(\kappa,\kappa_0) &= \frac{\sum_i}{N}\sum_{l=1}^{K}[E\{1(T_i \geq l)|X_i\}K[-2\hat{F}_Z\{Z_i(l)\} + 2F_Z\{Z_i(l)\} \\ &\quad +2\hat{F}_{Z_0}\{Z_{0,i}(l)\} - 2F_{Z_0}\{Z_{0,i}(l)\}]]. \qquad (11)\end{aligned}$$

Define

$$R_i(l) = E\{1(T_i \geq l)|X_i\}[-\hat{F}_Z\{Z_i(l)\} + F_Z\{Z_i(l)\} + \hat{F}_{Z_0}\{Z_{0,i}(l)\} - F_{Z_0}\{Z_{0,i}(l)\}]$$

so that

$$\triangle^A_{2,N}(\kappa,\kappa_0) = \frac{2K}{N}\sum_i\sum_{l=1}^{K} R_i(l).$$

Using $\hat{F}_Z\{Z_i(l)\} = \frac{\sum_{r\neq i}}{N-1}\frac{\sum_{k=1}^{K}}{K}1\{Z_r(k) < Z_i(l)\} + \frac{l-1}{(N-1)K}$ yields

$$R_i(l) = E\{1(T_i \geq l)|X_i\}\frac{\sum_{r\neq i}}{N-1}\frac{\sum_{k=1}^{K}}{K}[-1\{Z_r(k) < Z_i(l)\}+F_Z\{Z_i(l)\}+1\{Z_{0,r}(k) < Z_{0,i}(l)\}-F_{Z_0}\{Z_{0,i}(l)\}].$$

Note that $E\{R_i(l)\} = 0$ since that data is a random sample (i.i.d.). This gives

$$E\{\triangle^A_{2,N}(\kappa,\kappa_0)\} = 0 \text{ for all } \kappa,\kappa_0.$$

Next, note that

$$E([\hat{F}_Z\{Z_i(l)\} - F_Z\{Z_i(l)\}]^2) \leq \frac{1}{4}\frac{1}{N-1}$$

so that

$$E[\{R_i(l)\}^2] \leq \frac{1}{N-1}$$

for all i,l. Similarly, $\frac{-1}{N-1} \leq E[\{R_i(l)\}\{R_j(k)\}] \leq \frac{1}{N-1}$ for all l,k and $i \neq j$. This expression is helpful for evaluation the second moment of $\triangle^A_{2,N}(\kappa,\kappa_0)$. In particular,

$$\begin{aligned}E[\{\triangle^A_{2,N}(\kappa,\kappa_0)\}^2] &= (\frac{2K}{N})^2E[\{\sum_i\sum_{l=1}^{K}R_i(l)\}^2] \\ &= (\frac{2K}{N})^2E[\{\sum_i\sum_{l=1}^{K}R_i(l)\}\{\sum_j\sum_{k=1}^{K}R_j(k)\}].\end{aligned}$$

This gives

$$\begin{aligned}E[\{\triangle^A_{2,N}(\kappa,\kappa_0)\}^2] &= (\frac{2K}{N})^2E[\sum_i\{\sum_{l=1}^{K}R_i(l)\}^2] \\ &\quad +(\frac{2K}{N})^2E[\{\sum_i\sum_{l=1}^{K}R_i(l)\}\{\sum_{j\neq i}\sum_{k=1}^{K}R_j(k)\}].\end{aligned}$$

The first term is bounded by $\frac{4K^3}{N(N-1)}$ since $E[\{\sum_{l=1}^{K} R_i(l)\}^2]$ is bounded by $\frac{K}{N-1}$. Note that this bound is uniform, i.e. $\sup_{\kappa,\kappa_0\in\Theta} (\frac{2K}{N})^2 E[\sum_i \{\sum_{l=1}^{K} R_i(l)\}^2] < \frac{4K^3}{N(N-1)}$. Now consider the second term in the last equation. As shown above,

$$R_i(l) = E\{1(T_i \geq l)|X_i\}\frac{\sum_{r\neq i}}{N-1}\frac{\sum_{k=1}^{K}}{K}[-1\{Z_r(k) < Z_i(l)\}+F_Z\{Z_i(l)\}+1\{Z_{0,r}(k) < Z_{0,i}(l)\}-F_{Z_0}\{Z_{0,i}(l)\}].$$

Let $i \neq j$ and consider

$$E\{R_i(l)R_j(k)|X_i, X_j\} = E(E[\{1(T_i \geq l)|X_i\}E\{1(T_j \geq k)|X_j\}*$$

$$*[\frac{\sum_{r\neq i}}{N-1}\frac{\sum_{q=1}^{K}}{K}[-1\{Z_r(q) < Z_i(l)\} + F_Z\{Z_i(l)\} + 1\{Z_{0,r}(q) < Z_{0,i}(l)\} - F_{Z_0}\{Z_{0,i}(l)\}]$$

$$*\frac{\sum_{s\neq j}}{N-1}\frac{\sum_{q=1}^{K}}{K}[-1\{Z_s(q) < Z_j(k)\}+F_Z\{Z_j(k)\}+1\{Z_{0,s}(q) < Z_{0,j}(k)\}-F_{Z_0}\{Z_{0,j}(k)\}]]]|X_i, X_j).$$

Note that the *i.i.d.* assumption implies that the last term simplifies. In particular, if $r \neq s$, $r \neq i, j$, and $s \neq i, j$ then[23]

$$\begin{aligned} E([-1\{Z_r(k) \quad &< \quad Z_i(l)\} + F_Z\{Z_i(l)\} + 1\{Z_{0,r}(k) < Z_{0,i}(l)\} - F_{Z_0}\{Z_{0,i}(l)\}] \\ *[-1\{Z_s(k) \quad &< \quad Z_j(l)\} + F_Z\{Z_j(l)\} + 1\{Z_{0,s}(k) < Z_{0,j}(l)\} - F_{Z_0}\{Z_{0,j}(l)\}]|X_i, X_j) = 0. \end{aligned}$$

We first write $E\{R_i(l)R_j(k)|X_i, X_j\}$ using summations over $r \neq i, j$ and $s \neq i, j$. That is

$$E\{R_i(l)R_j(k)|X_i, X_j\} = E(E[\{1(T_i \geq l)|X_i\}E\{1(T_j \geq k)|X_j\}*$$

$$\begin{aligned} *([\frac{\sum_{r\neq i,j}}{N-1}\frac{\sum_{q=1}^{K}}{K} - 1\{Z_r(q) \quad &< \quad Z_i(l)\} + F_Z\{Z_i(l)\} + 1\{Z_{0,r}(q) < Z_{0,i}(l)\} - F_{Z_0}\{Z_{0,i}(l)\}] \\ &+\frac{\sum_{q=1}^{K}}{K}[-1\{Z_j(q) \quad < \quad Z_i(l)\} + F_Z\{Z_i(l)\} + 1\{Z_{0,j}(q) < Z_{0,i}(l)\} - F_{Z_0}\{Z_{0,i}(l)\}]) \end{aligned}$$

$$\begin{aligned} *[(\frac{\sum_{s\neq i,j}}{N-1}\frac{\sum_{q=1}^{K}}{K} - 1\{Z_s(q) \quad &< \quad Z_j(k)\} + F_Z\{Z_j(k)\} + 1\{Z_{0,s}(q) < Z_{0,j}(k)\} - F_{Z_0}\{Z_{0,j}(k)\}] \\ &+\frac{\sum_{q=1}^{K}}{K}[-1\{Z_i(q) \quad < \quad Z_j(k)\} + F_Z\{Z_J(k)\} + 1\{Z_{0,i}(q) < Z_{0,j}(k)\} - F_{Z_0}\{Z_{0,j}(k)\}])]|X_i, X_j). \end{aligned}$$

Removing the terms that are zero in expectation yields

$$\begin{aligned} E\{R_i(l)R_j(k)|X_i, X_j\} \quad &= \quad \frac{1}{N-1}E[E\{1(T_i \geq l)|X_i\}E\{1(T_j \geq k)|X_j\} * \\ *[\frac{\sum_{r\neq i,j}}{N-1}\frac{\sum_{q=1}^{K}}{K}[-1\{Z_r(q) \quad &< \quad Z_i(l)\} + F_Z\{Z_i(l)\} + 1\{Z_{0,r}(q) < Z_{0,i}(l)\} - F_{Z_0}\{Z_{0,i}(l)\}] \end{aligned}$$

[23] Let $r \neq i, j$ denote that $r \neq i$ and $r \neq j$.

$$*\frac{\sum_{q=1}^{K}}{K}[-1\{Z_r(q) < Z_j(k)\} + F_Z\{Z_j(k)\} + 1\{Z_{0,r}(q) < Z_{0,j}(k)\} - F_{Z_0}\{Z_{0,j}(k)\}]]|X_i, X_j]$$

Replacing q by $\tilde{q}$ in the last summation yields

$$\begin{aligned} E\{R_i(l)R_j(k)|X_i, X_j\} &= \frac{1}{N-1}E[E\{1(T_i \geq l)|X_i\}E\{1(T_j \geq k)|X_j\} * \\ *[\frac{\sum_{r\neq i,j}}{N-1}\frac{\sum_{q=1}^{K}}{K}\frac{\sum_{\tilde{q}=1}^{K}}{K}[-1\{Z_r(q) &< Z_i(l)\} + F_Z\{Z_i(l)\} + 1\{Z_{0,r}(q) < Z_{0,i}(l)\} - F_{Z_0}\{Z_{0,i}(l)\}] \\ *[-1\{Z_r(\tilde{q}) < Z_j(k)\} + F_Z\{Z_j(k)\} + 1\{Z_{0,r}(\tilde{q}) < Z_{0,j}(k)\} - F_{Z_0}\{Z_{0,j}(k)\}]]|X_i, X_j]. \end{aligned}$$

Define $F_Z\{Z_i(l), q\} = E[1\{Z_r(q) < Z_i(l)\}|X_i, X_j]$ for $r \neq i, j$ and $F_{Z_0}\{Z_{0,i}(l), q\} = E[1\{Z_{0,r}(q) < Z_{0,i}(l)\}|X_i, X_j]$ for $r \neq i, j$. Note that $F_Z\{Z_i(l)\} - F_{Z_0}\{Z_{0,i}(l)\} = \sum_{\tilde{q}=1}^{K} F_Z\{Z_i(l), q\} - \sum_{\tilde{q}=1}^{K} F_{Z_0}\{Z_{0,i}(l), q\}$. This gives

$$\begin{aligned} E\{R_i(l)R_j(k)|X_i, X_j\} &= \frac{1}{N-1}E[E\{1(T_i \geq l)|X_i\}E\{1(T_j \geq k)|X_j\} * \\ *[\frac{\sum_{r\neq i,j}}{N-1}\frac{\sum_{q=1}^{K}}{K}\frac{\sum_{\tilde{q}=1}^{K}}{K}[1\{Z_r(q) &< Z_i(l)\}1\{Z_r(\tilde{q}) < Z_j(k)\} - F_Z\{Z_i(l), q\}F_Z\{Z_j(k), \tilde{q}\} \\ -1\{Z_r(q) < Z_i(l)\}1\{Z_{0,r}(\tilde{q}) < Z_{0,j}(k)\} &- F_Z\{Z_i(l), q\}F_{Z_0}\{Z_{0,j}(k), \tilde{q}\} \\ -1\{Z_{0,r}(q) < Z_{0,i}(l)\}1\{Z_r(\tilde{q}) < Z_j(k)\} &- F_{Z_0}\{Z_{0i}(l), q\}F_Z\{Z_j(k), \tilde{q}\} \\ +1\{Z_{0,r}(q) < Z_{0,i}(l)\}1\{Z_{0,r}(\tilde{q}) < Z_{0,j}(k)\} &- F_{Z_0}\{Z_{0i}(l), q\}F_{Z_0}\{Z_{0,j}(k), \tilde{q}\}]]|X_i, X_j]. \end{aligned}$$

Next, define $\bar{\triangle}_{2,N}^{A}(\kappa, \kappa_0) = (\frac{2K}{N})^2[\sum_i \sum_{j\neq i} E[\{\sum_{l=1}^{K} R_i(l)\}\{\sum_{k=1}^{K} R_j(k)\}|X_i, X_j]$. Notice that $\bar{\triangle}_{2,N}^{A}(\kappa, \kappa_0)$ is $O(N^{-1})$ for any value of κ and κ_0. Moreover, notice that $\bar{\triangle}_{2,N}^{A}(\kappa, \kappa_0)$ is zero when evaluated at $\kappa = \kappa_0$ since $E\{R_i(l)R_j(k)|X_i, X_j\}|_{\kappa=\kappa_0} = 0$. Finally, $\bar{\triangle}_{2,N}^{A}(\kappa, \kappa_0)$ is differentiable with respect to κ since the expectation operator supplies the necessary smoothness. In particular,

$$\frac{\partial}{\partial\kappa}\bar{\triangle}_{2,N}^{A}(\kappa, \kappa_0) = \frac{\partial}{\partial\kappa}(\frac{2K}{N})^2 \sum_i \sum_{j\neq i} E[\{\sum_{l=1}^{K} R_i(l)\}\{\sum_{k=1}^{K} R_j(k)\}|X_i, X_j]$$

$$\begin{aligned} &= \frac{1}{N-1}(\frac{2K}{N})^2 E[\sum_i \sum_{j\neq i} E\{1(T_i \geq l)|X_i\}E\{1(T_j \geq k)|X_j\} * \\ &*[\frac{\sum_{r\neq i,j}}{N-1}\frac{\sum_{q=1}^{K}}{K}\frac{\sum_{\tilde{q}=1}^{K}}{K}[\{f_{rqil|\tilde{q}jk}(X_r) - f_{rqil}(X_r)\}F_Z\{Z_i(l), q\}[\frac{\partial}{\partial\kappa}\{Z_r(q, x) - Z_i(l)\}]|_{x=X_r} \\ &+\{f_{r\tilde{q}jk|qil}(X_r) - f_{r\tilde{q}jk}(X_r)\}F_Z\{Z_i(l), q\}[\frac{\partial}{\partial\kappa}\{Z_r(q, x) - Z_j(k)\}]|_{x=X_r} \\ &-\{f_{rqil|\tilde{q}jk}(X_r) - f_{rqil}(X_r)\}F_{Z_0}\{Z_{0,j}(k), \tilde{q}\}[\frac{\partial}{\partial\kappa}\{Z_r(q, x) - Z_i(l)\}]|_{x=X_r} \\ &-\{f_{r\tilde{q}jk|qil}(X_r) - f_{r\tilde{q}jk}(X_r)\}F_{Z_0}\{Z_{0,i}(l), q\}[\frac{\partial}{\partial\kappa}\{Z_r(q, x) - Z_j(k)\}]|_{x=X_r}]|X_i, X_j]]] \end{aligned} \quad (12)$$

where $f_{rqil|\tilde{q}jk}(.)$ is the density of X_r conditional on $Z_r(q) = Z_i(l)$ and $Z_r(\tilde{q}) < Z_j(k)$. Similarly, $f_{r\tilde{q}jk}(X_r)$ is the density of X_r conditional on $Z_r(q) = Z_i(l)$ (i.e. without the conditioning). The densities $f_{r\tilde{q}jk|qil}(.)$ and $f_{r\tilde{q}jk}(.)$ are similarly defined. Evaluating these densities at the random variable X_r gives a notation that is shorter than writing the expression as an integral. For clarity $Z_r(q,x)$ is explicitly written as a function of x since more than one value of x can yield the same value of $Z_r(q,x)$.

Notice that $\frac{\partial}{\partial\kappa}\bar{\triangle}^A_{2,N}(\kappa,\kappa_0)|_{\kappa=\kappa_0} = 0$ and $\frac{\partial}{\partial\kappa}E\{\bar{\triangle}^A_{2,N}(\kappa,\kappa_0)\}|_{\kappa=\kappa_0} = 0$. Thus, the first term in the Taylor expansion of $E\{\bar{\triangle}^A_{2,N}(\kappa,\kappa_0)\}$ around $\kappa = \kappa_0$ is zero. Also notice that $\bar{\triangle}^A_{2,N}(\kappa,\kappa_0)$ and $E\{\bar{\triangle}^A_{2,N}(\kappa,\kappa_0)\}$ are proportional to $\frac{1}{N-1}$ for any value of κ. Finally, notice that $\bar{\triangle}^A_{2,N}(\kappa,\kappa_0)$ and $E\{\bar{\triangle}^A_{2,N}(\kappa,\kappa_0)\}$ are twice differentiable with respect to κ so that we can use a Taylor expansion[24] with an error term that is proportional to $\frac{1}{N-1}||\kappa-\kappa_0||^2$. This implies that the asymptotic variation of

$$\frac{\sqrt{N}}{||\kappa-\kappa_0||^2}(\frac{2K}{N})^2E[\{\sum_i\sum_{l=1}^K R_i(l)\}\{\sum_{j\neq i}\sum_{k=1}^K R_j(k)\}].$$

is uniformly bounded for $\kappa,\kappa_0 \in \Theta$. This, together with our earlier result that $\sup_{\kappa,\kappa_0\in\Theta}(\frac{2K}{N})^2E[\sum_i\{\sum_{l=1}^K R_i(l)\}^2] < \frac{4K^3}{N(N-1)}$ implies that $\triangle^A_{2,N}(\kappa,\kappa_0)$ is $O_p(\frac{1}{N})+O_p(\frac{||\kappa-\kappa_0||}{\sqrt{N}})$ uniformly over $o_p(1)$ neighborhoods of κ_0. Thus, the conditions of Sherman (1993, theorem 1) are satisfied for $\triangle^A_{2,N}(\kappa,\kappa_0)$.

Next, consider $\triangle^B_{2,N}(\kappa,\kappa_0)$,

$$\triangle^B_{2,N}(\kappa,\kappa_0) = \frac{\sum_i}{N}\sum_{l=1}^K(E\{1(T_i\geq l)|X_i\}K[-2F_Z\{Z_i(l)\}+2F_{Z_0}\{Z_{0,i}(l)\}])$$
$$-\frac{\sum_i}{N}E\{\sum_{l=1}^K(E\{1(T_i\geq l)|X_i\}K[-2F_Z\{Z_i(l)\}+2F_{Z_0}\{Z_{0,i}(l)\}])\},$$

and note that $\triangle^B_{2,N}(\kappa,\kappa_0)$ is twice continuously differentiable. An expansion around κ_0 and using the fact that the expected value of the objective function is minimized at κ_0 yields that $\triangle^B_{2,N}(\kappa,\kappa_0)$ is $o_p(||\kappa-\kappa_0||^2)$. Therefore, $\triangle_{1,N}(\kappa,\kappa_0)+\triangle_{2,N}(\kappa,\kappa_0)$ is $O_p(||\kappa-\kappa_0||/\sqrt{N})+o_p(||\kappa-\kappa_0||^2)+O_p(1/N)$ uniformly over $o_p(1)$ neighborhoods of κ_0. Thus, Sherman (1993, theorem 1, condition (ii)) is satisfied so that $(\hat{\kappa}-\kappa_0)$ is $O_p(N^{-1/2})$.

[24] See for example Bronshtein and Semendyayev (1997, page 245).

Sherman's (1993) theorem 2 has three assumptions. The first one, that $(\hat{\kappa} - \kappa_0)$ is $O_p(1/\sqrt{N})$, was shown above. The second assumption, that κ is an interior point of Θ was assumed. The third assumption is that, uniformly over $O_p(1/\sqrt{N})$ neighborhoods of κ_0,

$$Q(\kappa) = Q(\kappa_0) - \frac{1}{2}(\kappa - \kappa_0)'V(\kappa - \kappa_0) + \frac{1}{\sqrt{N}}(\kappa - \kappa_0)'D_N(\kappa_0) + o_p(1/N)$$

where V is a negative definite matrix, and W_N converges in distribution to a $N(0, \Omega)$ random vector. The expected value of the objective function, $E\{Q(\kappa)\}$, is minimized at κ_0 and can be approximated by a second order Taylor expansion around κ_0. This approximation has the correct order of the error term since the neighborhoods are $O_p(1/\sqrt{N})$. Thus, we only need to show that

$$Q(\kappa) = Q(\kappa_0) + E\{Q(\kappa)\} - E\{Q(\kappa_0)\} + \frac{1}{\sqrt{N}}(\kappa - \kappa_0)'D_N(\kappa_0) + o_p(1/N). \tag{13}$$

Thus, we need to choose $\tilde{D}_N(\kappa)$ such that $\frac{1}{\sqrt{N}}(\kappa - \kappa_0)'\tilde{D}_N(\kappa_0)$ approximates $\triangle_{1,N}^{main}(\kappa, \kappa_0)$ and matches its variation. As before

$$\triangle_{1,N}^{main}(\kappa, \kappa_0) = \frac{\sum_i}{N}\sum_{l=1}^{K}[1(T_i \geq l) - E\{1(T_i \geq l)|X_i\}]K[-2F_Z\{Z_i(l)\} + 2F_{Z_0}\{Z_{0,i}(l)\}]$$

and define

$$\tilde{D}_N(\kappa) = \frac{1}{\sqrt{N}}\sum_i\sum_{l=1}^{K}[1(T_i \geq l) - E\{1(T_i \geq l)|X_i\}][\frac{\partial F_Z\{Z_i(l)\}}{\partial \kappa}|_{\kappa=\kappa_0}](-2K)$$

Choosing this $\tilde{D}_N(\kappa)$ yields the result ensures that equation (13) holds. In the main text, we divide $\tilde{D}_N(\kappa)$ by K, multiply by (-1) and define

$$D_N(\kappa) = \frac{2}{\sqrt{N}}\sum_i\sum_{l=1}^{K}[1(T_i \geq l) - E\{1(T_i \geq l)|X_i\}][\frac{\partial F_Z\{Z_i(l)\}}{\partial \kappa}|_{\kappa=\kappa_0}].$$

Changing the sign has no effect on the asymptotic distribution and dividing by K is just a normalization (K still affects $D_N(\kappa)$ through the distribution of Z). The assumption that $F_Z\{Z_i(l;\kappa)$ is twice continuously differentiable in a neighborhood $\mathcal{N}$ of κ_0 for any l, the random sample assumption of Assumption 1, and the Lindeberg-Levy central limit theorem imply that $D_N(\kappa)$ converges to a normal distribution with variance-covariance $\Omega = E[D_N(\kappa_0)D_N(\kappa_0)']$. Matrix multiplication then implies the asymptotic variance and the result follows.

Q.E.D.

Appendix 4:

Proof of **Proposition 1**

Horowitz (2001, Theorem 2.2) shows that bootstrapping an asymptotically normally distributed estimator that can be represented by an influence function yields a consistent variance-covariance matrix.[25] Thus, $\hat{V}_\kappa = H^{-1}\Omega H^{-1} + o_p(1)$ where $H^{-1}\Omega H^{-1}$ is the asymptotic variance-covariance matrix derived in theorem 2. Hausman's (1978) proof compares parametric estimators. The same proof applies here since the fact that the estimator $\hat{\kappa}$ allows for a nonparametric unobserved heterogeneity distribution does not play a role. In particular, we can take the properties of $\hat{\omega}$ and $\hat{\kappa}$ as primitives and the result follows.

Appendix: 5 Counterexample to Han (1987)

The examples below show that the conditions of Han (Journal of Econometrics, 1987) are not sufficient for identification of the MRC estimator. Consider the following data generating processes. In both models, the baseline hazard does *not* depend on time.

Model I:

Let X_1 be distributed as a standard normal or another distribution with support on the whole real line. Let $X_2 = \ln(-\ln[\frac{\exp\{-\frac{1}{2}\exp(X_1)\}+\exp\{-\frac{3}{2}\cdot\exp(X_1)\}}{2}])$. Note that $\frac{\exp\{-\frac{1}{2}\exp(X_1)\}+\exp\{-\frac{3}{2}\cdot\exp(X_1)\}}{2}$ has support on (0,1), $-\ln[\frac{\exp\{-\frac{1}{2}\exp(X_1)\}+\exp\{-\frac{3}{2}\cdot\exp(X_1)\}}{2}]$ has support on the positive real line and X_2 has support on the whole real line. Let

$$\theta(t|X, v) = v \cdot \exp(X_1),$$

and let

$$p(v = \frac{1}{2}|X) = p(v = \frac{3}{2}|X) = \frac{1}{2}.$$

[25] Horowitz (2001, Theorem 2.2) averages $g_n(X_i)$.

This gives the following probability of survival for the first period:

$$S(t=1|X) = \frac{\exp\{-\frac{1}{2}\exp(X_1)\} + \exp\{-\frac{3}{2}\cdot\exp(X_1)\}}{2}.$$

Model II:

Let X_1, X_2 be the same as above. Let $\theta(t|X,v) = v\cdot\exp(X_2)$, and let $p(v=1|X)=1$. This gives the following probability of survival for the first period:

$$S(t=1|X) = \frac{\exp\{-\frac{1}{2}\exp(X_1)\} + \exp\{-\frac{3}{2}\cdot\exp(X_1)\}}{2}.$$

Note that Model I and Model II yield the same probability of survival for the first period. Suppose that we only observe whether or not individuals survive the first period. Then, Model I and Model II are observationally equivalent and identification or estimation is not possible. All the assumptions of Han (1987) are satisfied. Note that the model is not identified in the sense that there are two values of β that yield the same density and that $||\beta|| = 1$ in each case. Thus, stronger assumptions are needed. Assuming that $\beta_1 \neq 0$ is sufficient if all assumptions of Han (1987) are maintained.

Appendix: 6 Examples

Example 1:

Consider the following hazard model: $\theta(t|v,X) = \phi^X\lambda(t)$ so that $\bar{F}(t \mid X) = \exp(-\phi^X\Lambda(t))$ and (ii) $\bar{F}(t \mid X)$ be observed for $X = 0, 1$ and all $t \geq 0$.

We first estimate the integrated baseline hazard, $\hat{\Lambda}(t) = -\ln\{\bar{F}(t \mid X=0)\} = -\ln\{\bar{F}_{X=0}(t)\}$. This implies the following density: $f(t \mid X=1) = \phi\lambda(t)e^{-\phi\Lambda(t)}$. After estimating $\lambda(t)$ and $\Lambda(t)$ nonparametrically using $\bar{F}_{X=0}(t)$, we can derive the log likelihood by conditioning on these estimated functions (as in Newey and McFadden (1994). This yields,

$$\begin{aligned} L(\phi) &= \ln\phi^X + \ln\lambda(t) - \phi^X\Lambda(t) \\ \frac{\partial L(\phi)}{\partial\phi} &= \frac{X}{\phi} - X\Lambda(t) \Rightarrow \\ \operatorname*{plim}_{N\to\infty}\hat{\phi}_{MLE} &= 1/E\{\Lambda(T)|X=1\} = -1/E[\ln\{\bar{F}_0(T)\}|X=1]. \end{aligned}$$

Let the data be generated by the following model $\theta(t|v,X) = v\phi^X$ where $v \sim \text{Gamma}(\alpha,\alpha)$. Thus, $\bar{F}_0(t) = \left(1+\frac{\Lambda(t)}{\alpha}\right)^{-\alpha}$ and $-\ln F_0(t) = \alpha \ln\left(1+\frac{\Lambda(t)}{\alpha}\right)$. Note that $\phi^X\Lambda(T) = \frac{Z}{v}$, where Z has an exponential distribution with mean one. This yields

$$\operatorname*{plim}_{N\to\infty}\hat{\phi} = \frac{1}{E[\alpha\ln\{1+Z/(\phi v\alpha\}]},$$

where $v \sim \text{Gamma}(\alpha,\alpha)$. Note that ϕ only appears in the denominator of the argument of a logarithmic function. Simulation gave the numerical result.

Example 2:

After transforming the dependent variable using the transformation model of Horowitz (1996), we define $W = H(T)$. Note that $H(T)^{|\beta|}$ is distributed as an exponential random variable, so W is distributed as a Weibull random variable with parameter $|\beta|$. As in the example, let $\beta > 0$. Consider the Weibull model with a Gamma mixing distribution, which is given by

$$\begin{aligned}
\theta\left(w_i \mid v, X_i\right) &= v e^{X_i\beta}\alpha w_i^{\alpha-1} \\
v &\sim \text{Gamma}\left(\gamma_v,\delta_v\right) \\
\bar{F}(w_i|X_i) &= E v e^{-v e^{X_i\beta}t_i^\alpha} = \frac{1}{\left(1+\frac{e^{X_i\beta}w_i^\alpha}{\delta_v}\right)^{\gamma_v}} \\
f(w_i|X_i) &= \frac{\alpha\gamma_v e^{X_i\beta}w_i^\alpha}{\delta_v}\frac{1}{\left(1+\frac{e^{X_i\beta}w_i^\alpha}{\delta_v}\right)^{\gamma_v+1}} \\
L^i(\alpha,\beta,\gamma_v,\delta_v) &= \ln\alpha + \ln\gamma_v + X_i\beta + \alpha\ln W_i - \ln\delta_v - (\gamma_v+1)\ln\left(1+\frac{e^{X_i\beta}W_i^\alpha}{\delta_v}\right).
\end{aligned}$$

Imposing the restriction $\alpha = \beta$ and using

$$e^{X_i\beta}W_i^\beta = (e^{X_i\beta_0}W_i^{\beta_0})^{\beta/\beta_0} = (Z_i)^{\beta/\beta_0},$$

where Z_i is distributed as an exponential random variable with mean one, gives

$$L^i(\beta,\gamma_v,\delta_v) = \ln\beta + \ln\gamma_v + (\beta/\beta_0)\ln Z_i - \ln\delta_v - (\gamma_v+1)\ln\left(1+\frac{(Z_i)^{\beta/\beta_0}}{\delta_v}\right).$$

This likelihood does not depend on the regressor[26] x, which implies that the probability limit of β does not depend on the distribution of X.

[26] The same reasoning holds for a negative β_0 (since the sign can be determined using Han (1987)) and for a multivariate regressor (since this can be reduced to a scalar by estimating the regression coefficient up to scale using Han (1987)); Han's estimator converges under the assumptions of the model. See the discussion above.

Appendix: 7 Computational Issues

by Matthew Harding, Jerry Hausman, and Tiemen Woutersen

We estimate the parameter vector (β, δ) from the following objective function which corresponds to a mass of indicator functions:

$$Q(\beta,\delta) = \sum_{i=1}^{N}\sum_{l=1}^{K} 1\{T_i \geq l\} \sum_{j=1}^{N}\sum_{k=1}^{K} [1\{Z_i(l) < Z_j(k)\} - 1\{Z_i(l) > Z_j(k)\}]. \tag{14}$$

Optimization of this objective function using iterated sums is not feasible because for the specification with 24 periods, it takes approximately 15 minutes to evaluate one such objective function in Matlab. Note, however, that for all individuals i who pass the criterion $T_i \geq l$, the objective function evaluates the difference between the number of individuals with an index less than the index of individual i and the number of individuals with an index greater than the index of individual i. This information is also contained in the ranking of individuals' indices and thus can be more efficiently extracted using the rank function. This suggests that an efficient implementation of this optimization will be similar to that of Chen (2002).

We can define $d_k = 1\{T \geq k\}$ for the vector T of dimension $N \times 1$. Let d be constructed by stacking the vectors d_k vertically for all $k = 1, ..., K$. Similarly let Z be constructed by stacking the vectors $Z(k)$ for all $k = 1, ..., K$. Now both d and Z are of dimension $NK \times 1$. We can now rewrite $Q(\beta, \delta)$ using these vectors and the rank function:

$$Q(\beta,\delta) = \frac{1}{N(N-1)} \sum_{i=1}^{NK} d(i) \left[2 \cdot Rank(Z(i)) - NK\right]. \tag{15}$$

This simpler yet numerically identical representation[27] will be more efficient to evaluate numerically because (i) it has only one summation sign and (ii) computation of the rank function requires sorting for which highly efficient algorithms are available. Indeed it now takes less than one second to estimate one such objective function for the specification with 24 periods.

[27] There is still an issue regarding the treatment of ties in the rank function but it seems to matter little in practice.

Models with non-smooth objective functions in the parameters have been traditionally estimated using the Nelder-Mead simplex method (Abrevaya (1999); Cavanagh and Sherman (1998)). In this particular example, the large number of local optima makes the Nelder-Mead method computationally unstable. The Nelder-Mead algorithm fails to converge or takes unreasonably long to do so.[28]

Pattern search methods have been available for many decades and rigorous convergence results have become available in recent years (Lewis and Torczon (1999); Audet and Dennis (2003)). Although anecdotal evidence on the performance of these algorithms often suggests slow convergence, we find that the convergence of the objective function at 4 decimal places for the specification with 13 periods takes about 20 minutes, while the specification with 24 periods takes approximately 50 minutes to convergence.

We now provide a brief introduction to the mechanism of pattern search.[29] For some given real-valued objective function $Q(\gamma)$ defined on the n-dimensional Euclidean space, let γ_0 be the initial guess. In our case, we use $\gamma_0 = [\,\widehat{\beta}, \widehat{\delta}]_{Gamma}$, the parameter estimates from the HHM Gamma Heterogeneity model estimated using a quasi-Newton derivative-based method. Additionally, define a *forcing function* $\rho(t)$ to be a continuous function such that $\rho(t)/t \to 0$ as t$\to$ 0. Let Δ_k control the step length at each iteration.

Search patterns for some initial starting value γ_0 are drawn from a given *generating set.* A minimal generating set corresponds to some positive spanning set for the n-dimensional space, where the number of dimensions corresponds to the number of parameters to be estimated. The defining requirement for a generating set is that any vector in $\mathbb{R}^n$ may be written as a linear combination of elements in the generating set using positive coefficients only. A generating set will thus contain at least n+1 elements. To illustrate, the generating set for $n = 2$ is

$$G = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\}. \tag{16}$$

Alternatively, we could use the set of $2n$ coordinate directions as the elements of our generating set. In our application, however, we find computational performance is

[28] Convergence of the objective function to 4 decimal places may take as long as 9 hours to compute.
[29] For a more detailed review and convergence proofs, see Kolda, Lewis and Torczon (2003).

superior under the setup with $n+1$ directions. Additionally, heuristic additions to the generating set may be implemented in order to improve speed and performance. These heuristic additions allow the algorithm to evaluate other points in the same direction as the last successful search, but further away from the starting point than permitted by the standard elements of the generating set. This allows for the possibility that if the correct direction of improvement is found, several computation steps will be skipped and the search converges more rapidly. Random polling vectors also provide heuristic evaluations of the objective function without compromising the convergence properties of the algorithm, which only depend on the minimal generating set.

We use the standard errors of the HHM estimation to construct a "bounding box" that is then used to bound the parameter space for the optimization under the semi-parametric setup. We start with a bounding[30] box of $\pm$ 3 standard errors.

At each iteration, the algorithm evaluates the objective function for all vectors $g_k \in G$ and compares $Q(\gamma_k + \Delta_k g_k)$ with $Q(\gamma_k) - \rho(\Delta_k)$. If an improvement is found, $\gamma_{k+1} = \gamma_k + \Delta_k g_k$ and Δ_k is increased to Δ_{k+1}. If no improvement is found, then $\gamma_{k+1} = \gamma_k$ and Δ_k is decreased to Δ_{k+1}. This process is iterated to convergence.

Since the true parameter values are not guaranteed to lie within this bounding box, it may be that the algorithm constrained by the location and size of the bounding box only reaches a local optimum. In order to correct for this possibility, we gradually expand the bounding box as long as the estimated parameters change with a larger bounding box. A large bounding box, however, may imply that the estimates have only low precision, since the algorithm visits every point in the domain with a probability decreasing in the size of the bounding box. In order to improve accuracy, once the desired size of the bounding box has been reached, the bounding box is re-centered on the new parameter estimates from the semi-parametric setup. The size of the bounding box is then sequentially decreased in order to verify the accuracy of the obtained estimates. Refinements are made if an improvement is found.

We use the estimated values $\widehat{\delta}_{Pattern}$ to compute an estimate of the survival probability at each period. Using the delta method, we compute the associated estimates of the

[30] We would increase the number of standard errors if the sample size was larger.

standard error of the survival probability in each period. Interpretation is made easier by smoothing the pair $(P(T \geq t_i), t_i)$ for all time periods t_i using a local polynomial method. The neighborhood of t_i is defined as a percentage of the total number of periods under consideration and may be chosen using cross-validation techniques. Each point in the neighborhood $N(t_i)$ is assigned two sets of weights. One set of weights is inversely proportional to the standard error of the survivor estimate, as given by the pattern search optimization. The other set of weights is provided by the *tri-cubic weight function* and weighs the impact of distant data points on the smoothing estimate of one particular observation. The tri-cubic weight function involved in the smoothing of point t_i places the following weight on observation t_j:

$$W(t_i, t_j) = \left(1 - \left(\frac{|t_i - t_j|}{\max_{t_j \in N(t_i)} |t_i - t_j|}\right)^3\right)^3 \mathbf{1}\left\{0 \leq \frac{|t_i - t_j|}{\max_{t_j \in N(t_i)} |t_i - t_j|} < 1\right\}. \quad (17)$$

The smoothed estimates of the survivor function are then computed as the predicted values of the weighted linear regression of second degree for each point in the corresponding neighborhood using the two sets of weights. The choice of the span of the neighborhood at each point using cross-validation tends to matter little in this case.

The pattern search method we employ to derive estimates of the model parameters seems to perform well, both in terms of accuracy and computational time. Nevertheless, the nature of the objective function and the dependency of our use of the pattern search method on a good estimate of the relevant bounding box raises the question to what extent a global optimum has been reached for our objective function. Since it is possible to conceive of our optimization problem as a stochastic optimization problem, we consider the implementation of a *genetic optimization* procedure as a global optimizer capable of overcoming the nondifferentiability of the objective function, as discussed by Spall (2003). Few applications of this procedure to econometrics exist in spite of numerous reported successful implementations in other areas of science (Haupt and Haupt (1998); Reeves and Rowe (2003)).

Genetic optimization methods describe a number of processes based on principles from biological sciences aimed at generating a population of parameter values which optimizes its *fitness*, defined as the corresponding value of the objective function. The core idea

involves the use of stochastic perturbations in the population of potential optimizing parameters so as to improve the optimality of the solution. This approach mirrors the biological concept of evolution. The use of a population of parameters as the primary building block of the algorithm aims at avoiding convergence towards one local optimum.

Since the outcome of a genetic optimization procedure is not dependent on the initial population, we use as starting values for the population unit-uniform random numbers. The objective function is evaluated for each member of the population. Members of the population with the best values are selected as candidates for the generation of individuals of the subsequent population through the processes of elitism, crossover or mutation. A (small) number of the successful members of a population are simply copied over in the next generation of the population, a process termed elitism. The crossover process randomly combines values of the parameter vector of two evolutionary successful individuals to obtain a new individual for the next population. The process of mutation adds random noise from a normal distribution to the parameter values of one successful individual to create a new individual in the next generation. Since with each additional generation we are more likely to close-in on the optimum, we shrink the variance of the mutation process at each generation.

The genetic optimization process tends to converge much slower than the pattern search procedure. Nevertheless, the algorithm can be used to confirm the global optimality of the point estimates obtained by pattern search. Our results using genetic optimization are the same as with the pattern search algorithm to 4 significant digits for the objective function.

References

[1] Abrevaya, J. (1999): "Computation of the maximum rank correlation estimator," *Economics Letters* 62, 279–285.

[2] Audet, C. and J. E. Dennis, Jr. (2003): "Pattern Search Algorithms for Mixed Variable Programming," *SIAM Journal on Optimization* 11: 573-594.

[3] Baker, M. and A. Melino (2000): "Duration Dependence and Nonparametric Heterogeneity: A Monte Carlo Study," *Journal of Econometrics*, 96, 357-93.

[4] Bijwaard, G. and G. Ridder (2002): "Efficient Estimation of the Semi-parametric Mixed Proportional Hazard Model," in preparation.

[5] Bijwaard, G. and G. Ridder (2009): "A Simple GMM Estimator for the Semi-parametric Mixed Proportional Hazard," in preparation.

[6] Brinch, C. N. (2007): Nonparametric Identification of The Mixed Hazards Model With Time-Varying Covariates, *Econometric Theory*, 23: 349-354.

[7] Bronshtein, I. N. and K. A. Semendyayev (1997): *Handbook of Mathematics*, Springer Verlag, New York.

[8] Cavanagh, C., R. P. Sherman (1998): "Rank Estimators for monotonic index models," *Journal of Econometrics*, 84, 351-381.

[9] Chen, S. (2002): "Rank Estimation of Transformation Models," *Econometrica*, 70, 1683-96.

[10] Cox, D. R. (1972): "Regression models and life tables (with discussion)," *Journal of the Royal Statistical Society* B, 34: 187-220.

[11] 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.

[12] Frederiksen, A., B. E. Honoré, and L. Hu (2007): "Discrete time duration models with group-level heterogeneity," *Journal of Econometrics*, 141, 1014-1043.

[13] Hahn, J. (1994): "The Efficiency Bound of the Mixed Proportional Hazard Model, " *Review of Economic Studies*, 61, 607-629.

[14] Ham, J. C., and R. J. LaLonde (1996): "The Effect of Sample Selection and Initial Conditions in Duration Models; Evidence from Experimental Data on Training," *Econometrica*, 64, 175-205.

[15] Ham, J. C. and S. A. Rea (1987): "Unemployment Insurance and Male Unemployment Duration in Canada." *Journal of Labor Economics*, Vol. 5 (3), 325-353

[16] Han, A. K. (1987): "Non-parametric Analysis of a Generalized Regression Model, the Maximum Rank Correlation Estimator," *Journal of Econometrics,* 35, 303-316.

[17] Han, A. K. and J. A. Hausman (1990): "Flexible Parametric Estimation of Duration and Competing Risk Models," *Journal of Applied Econometrics.*

[18] Haupt, R.L. and S.E. Haupt (1998) *Practical Genetic Algorithms,* Wiley-Interscience.

[19] Hausman, J. A. (1978): "Specification Tests in Econometrics," *Econometrica*, 46, 1251-72.

[20] Hausman, J. A., and D. A. Wise (1979): "Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment," *Econometrica,* 47, 455-474.

[21] Hausman, J. A., and W. E. Taylor (1981): "Panel Data and Unobservable Individual Effects," *Econometrica,* 49, 1377-1398.

[22] Heckman, J. J. (1978): "Simple statistical models for discrete panel data developed and applied to test the hypothesis of true state dependence against the hypothesis of spurious state dependence," *Annales de l'insee*, 30-31, 227-269.

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

[24] Heckman, J. J., and B. Singer (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52, 271-320.

[25] Honoré, B. E. (1990): "Simple Estimation of a Duration Model with Unobserved Heterogeneity," *Econometrica*, 58, 453-473.

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

[27] Honoré, B. E. (1993b): "Identification Results for Duration Models with Multiple Spells or Time-Varying Regressors," Northwestern working paper.

[28] Honoré, B. E., and A. de Paula (2008): "Interdependent Durations," Princeton mimeo.

[29] Honoré, B. E., and L. Hu (2004): "Estimation of Cross Sectional and Panel Data Censored Regression Models with Endogeneity," *Journal of Econometrics*, 122, 293–316.

[30] Honoré, B. E., and L. Hu (2010): "Estimation of a transformation model with truncation, interval observation and time-varying covariates," *Econometric Journal*, 13, 127–144.

[31] Horowitz, J. L. (1996): "Semiparametric Estimation of a Regression Model with an Unknown Transformation of the Dependent Variable," *Econometrica*, 64, 103-107.

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

[33] Horowitz, J. L. (2001): "The Bootstrap" in *Handbook of Econometrics,* Vol. 5, ed. by J. J. Heckman and E. Leamer. Amsterdam: North-Holland.

[34] Horowitz, J. L. and S. Lee (2004): "Semiparametric estimation of a panel data proportional hazards model with fixed effects," *Journal of Econometrics*, 119, 155-198.

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

[36] Ishwaran, H. (1996b): "Uniform Rates of Estimation in the Semiparametric Weibull Mixture Model," *The Annals of Statistics*, 24, 1572-1585.

[37] Khan, S. and E. Tamer (2007): Partial Rank Estimation of Duration Models with General Forms of Censoring, *Journal of Econometrics*, 25, 251-280.

[38] Kendall, M. G. (1938): "A new measure for rank correlation," *Biometrica*, 30, 81-93.

[39] Kiefer, J. and J. Wolfowitz (1956): "Consistency of Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters", *Annals of Mathematical Statistics,* 27, 887-906.

[40] Kolda, T. G., Lewis, R. M., and Torczon, V. (2003): "Optimization by direct search: New perspectives on some classical and modern methods," *SIAM Review*, 45: 383–482.

[41] Lancaster, T. (1979): "Econometric Methods for the Duration of Unemployment," *Econometrica*, 47, 939-956.

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

[43] Lancaster, T. and S. J. Nickell, (1980): "The Analysis of Re-employment Probabilities for the Unemployed," *Journal of the Royal Statistical Society*, A, 143, 141-165.

[44] Lewis, R. M. and V. Torczon (1999): "Pattern search algorithms for bound constrained minimization," *SIAM Journal on Optimization*, 9: 1082–1099.

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

[46] Mundlak, Y. (1961): "Empirical Production Function Free of Management Bias," *Journal of Farm Economics,* 43, 44-56.

[47] Newey, W. K. (1991): "Uniform Convergence in Probability and Stochastic Equicontinuity," *Econometrica*, 59, 1161-1167.

[48] Newey, W. K., and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics,* Vol. 4, ed. by R. F. Engle and D. MacFadden. Amsterdam: North-Holland.

[49] Nielsen, G.G., Gill, R.D., Andersen, P.K. & Sørensen, T.I.A. (1992): "A Counting Process approach to maximum likelihood estimation in frailty models," *Scandanavian Journal of Statistics.* 19, 25-43

[50] Reeves, C.R. and J.E. Rowe (2003) *Genetic Algorithms - Principles and Perspectives: A Guide to GA Theory*, Kluwer Academic.

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

[52] Ridder, G. and T. Woutersen (2003): "The Singularity of the Information Matrix of the Mixed Proportional Hazard Model," *Econometrica*, 71, 1579-1589.

[53] Sherman, R. P. (1993): "The Limiting Distribution of the Maximum Rank Correlation Estimator," *Econometrica*, 61, 123-137.

[54] Spall, J. (2003) *Introduction to Stochastic Search and Optimization: Estimation, Simulation and Control*, Wiley-Interscience.

[55] Van den Berg, G. J. (2001): "Duration Models: Specification, Identification, and Multiple Duration," in *Handbook of Econometrics,* Vol. 5, ed. by J. J. Heckman and E. Leamer. Amsterdam: North-Holland.

[56] Van der Vaart, A.W. (1998): *Asymptotic Statistics*, Cambridge Series in Statistical and Probilistic Mathematics, Cambridge: Cambridge University Press.

[57] Woutersen, T. (2000): "Estimators for Panel Duration Data with Endogenous Censoring and Endogenous Regressors," UWO working paper.

[58] Woutersen, T. (2002): "Robustness Against Incidental Parameters," UWO working paper.

Table 1: Data Description and Summary Statistics

Variable	*Description*	*Mean*	*Standard Deviation*
Race	= 1 if UI recipient is Black or African-American	0.1172	0.3217
Age	= 1 if UI recipient is over 50 years old at the start of the benefit year	0.1776	0.3822
Replacement Rate	= Weakly Benefit Amount divided by UI recipient's base period earnings	0.0129	0.0076
State Unemployment Rate	= Unemployment rate of the state from which the individual received UI benefits during the period in which the individual filed for benefits		
	Week 1	4.6863	1.0875
	Week 2	4.6726	1.0834
	Week 3	4.6603	1.0794
	Week 4	4.6453	1.0747
	Week 5	4.6301	1.0698
	Week 6	4.6211	1.0649
	Week 7	4.6164	1.0665
	Week 8	4.5981	1.0641
	Week 9	4.5710	1.0616
	Week 10	4.5382	1.0615
	Week 11	4.5318	1.0630
	Week 12	4.5091	1.0678
	Week 13	4.4832	1.0751
	Week 14	4.4620	1.0802
	Week 15	4.4604	1.0756
	Week 16	4.4490	1.0735
	Week 17	4.4400	1.0675
	Week 18	4.4407	1.0557
	Week 19	4.4316	1.0546
	Week 20	4.4207	1.0452
	Week 21	4.4240	1.0337
	Week 22	4.4315	1.0298
	Week 23	4.4364	1.0240
	Week 24	4.4414	1.0156
	Week 25	4.4424	1.0121

Table 2: HHM Gamma Heterogeneity Model, Period 1 normalized to zero.

		6 periods		13 periods		24 periods	
		Parameters	**s.e.**	**Parameters**	**s.e.**	**Parameters**	**s.e.**
alpha		0.9307	2.1675	0.1089	0.0120	0.0993	0.0182
gamma		7.9607	0.2383	0.3164	0.0773	0.1655	0.6082
State Unemployment Rate		-0.1019	0.0246	-0.2762	0.0341	-0.3875	0.0393
Race		-0.0350	0.0653	-0.2167	0.1155	-0.2061	0.1370
Age>50		-0.2047	0.0623	-0.4290	0.0932	-0.4317	0.1557
Replacement Rate		-0.5393	0.0497	-0.5498	0.0562	-0.5059	0.1493
Period	2	-0.3259	0.0747	-0.0494	0.0787	0.0010	0.1576
	3	0.0198	0.0814	0.5517	0.0905	0.6479	0.1342
	4	-0.3032	0.0939	0.4661	0.1157	0.6053	0.1222
	5	0.1430	0.1026	1.1678	0.1275	1.3511	0.1532
	6	-0.3780	0.1256	0.8858	0.1553	1.1134	0.1979
	7			1.4905	0.1811	1.7608	0.1879
	8			1.3001	0.2086	1.6111	0.2144
	9			1.7490	0.2228	2.0944	0.2359
	10			1.7326	0.2486	2.1103	0.2753
	11			2.2152	0.2661	2.6362	0.3007
	12			2.3336	0.2870	2.7970	0.3510
	13			2.6485	0.3108	3.1545	0.3966
	14					3.4413	0.3856
	15					3.8034	0.4204
	16					3.7589	0.5024
	17					4.3672	0.5399
	18					4.4417	0.5073
	19					4.9485	0.5167
	20					4.9909	0.5785
	21					5.3740	0.5845
	22					5.4392	0.6022
	23					5.9363	0.6546
Period	24					6.0436	0.6891
Number of observations		**15,491**		**15,491**		**15,491**	
Likelihood		**0.6664**		**1.2242**		**1.0131**	

Table 3: New Duration Model, Period 1 normalized to zero.

		6 periods		13 periods		24 periods	
		Parameters	**s.e.**	**Parameters**	**s.e.**	**Parameters**	**s.e.**
State Unemployment Rate		-1.4672	0.0965	-1.4643	0.0832	-1.3953	0.0483
Race		-0.5663	0.2728	-0.5928	0.2444	-0.5656	0.2105
Age>50		-1.0701	0.2146	-1.0712	0.1974	-0.8067	0.1770
Replacement Rate		-2.2347	0.1778	-2.2693	0.1588	-0.5372	0.1097
Period	2	2.7287	0.1295	2.6191	0.1604	2.0707	0.2422
	3	3.8869	0.1298	4.1002	0.1812	3.2261	0.2451
	4	5.0912	0.1276	5.4381	0.1657	4.2821	0.2116
	5	5.6051	0.1440	5.9834	0.1737	4.7376	0.2132
	6	6.5985	0.1380	7.1400	0.1704	5.7784	0.2028
	7			7.1200	0.2092	5.6905	0.2444
	8			7.9306	0.1860	6.5007	0.1955
	9			8.2543	0.2017	6.7297	0.2212
	10			8.3960	0.2382	6.5937	0.3050
	11			8.7536	0.2265	7.1753	0.2422
	12			9.4656	0.2218	7.8302	0.2218
	13			9.7804	0.2361	8.3342	0.2227
	14					8.1757	0.3352
	15					8.4889	0.3058
	16					9.1671	0.2548
	17					9.5479	0.2597
	18					9.8108	0.2818
	19					10.0790	0.2968
	20					10.6790	0.3018
	21					10.7060	0.3229
	22					10.9360	0.3409
	23					10.9230	0.3419
Period	24					11.3860	0.3437
Number of observations		**15,491**		**15,491**		**15,491**	
Objective Function		**30.221**		**122.050**		**332.890**	

Table 4, Variance local unemployment rates

	Periods		
Standard D	6	13	24
Overall	1.6249	1.3728	1.2386
Between	0.8977	0.9519	0.9543
Within	1.3545	0.9892	0.7896

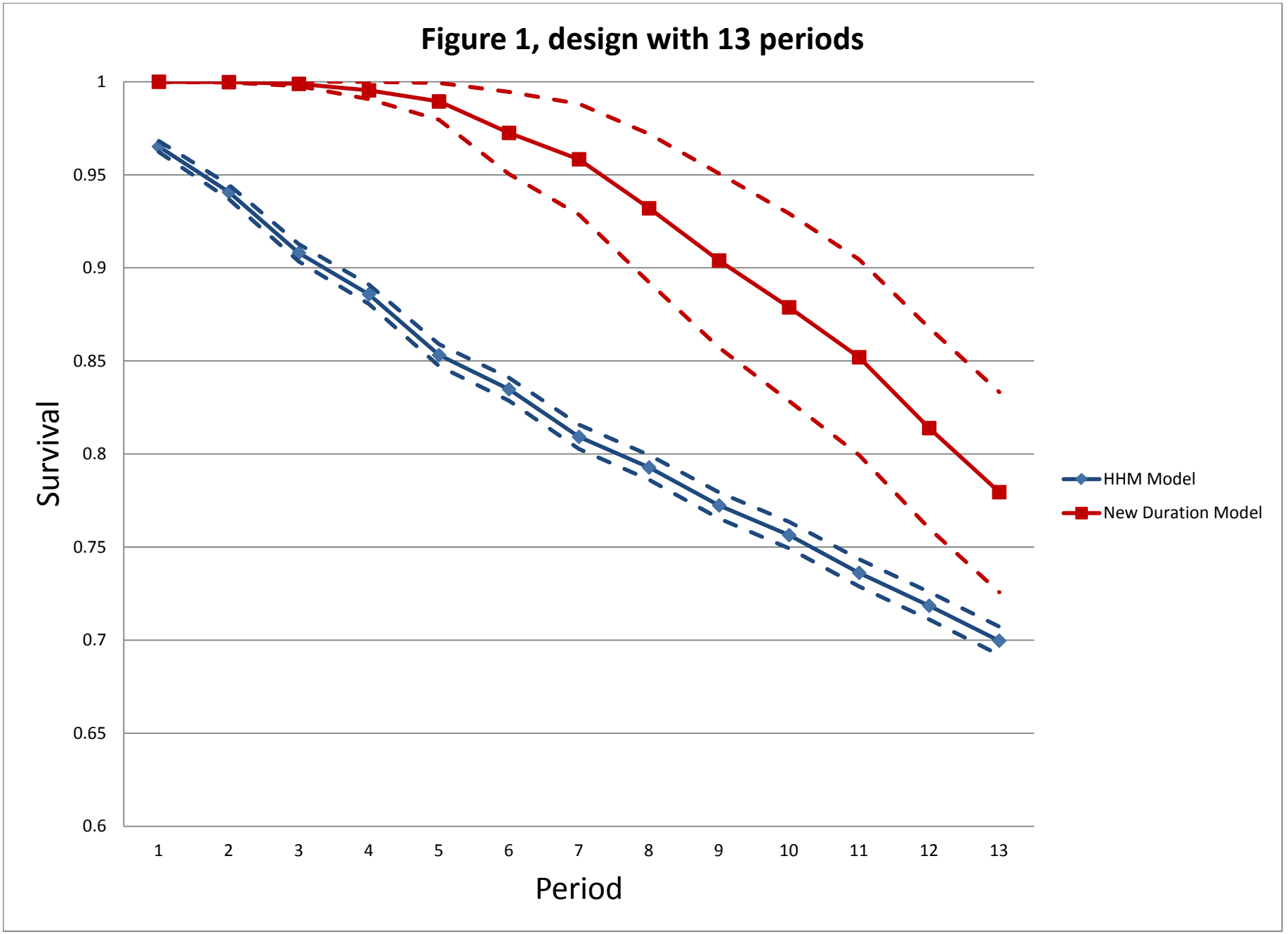
Figure 1, design with 13 periods
Survival
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
1
2
3
4
5
6
7
8
9
10
11
12
13
Period
HHM Model
New Duration Model

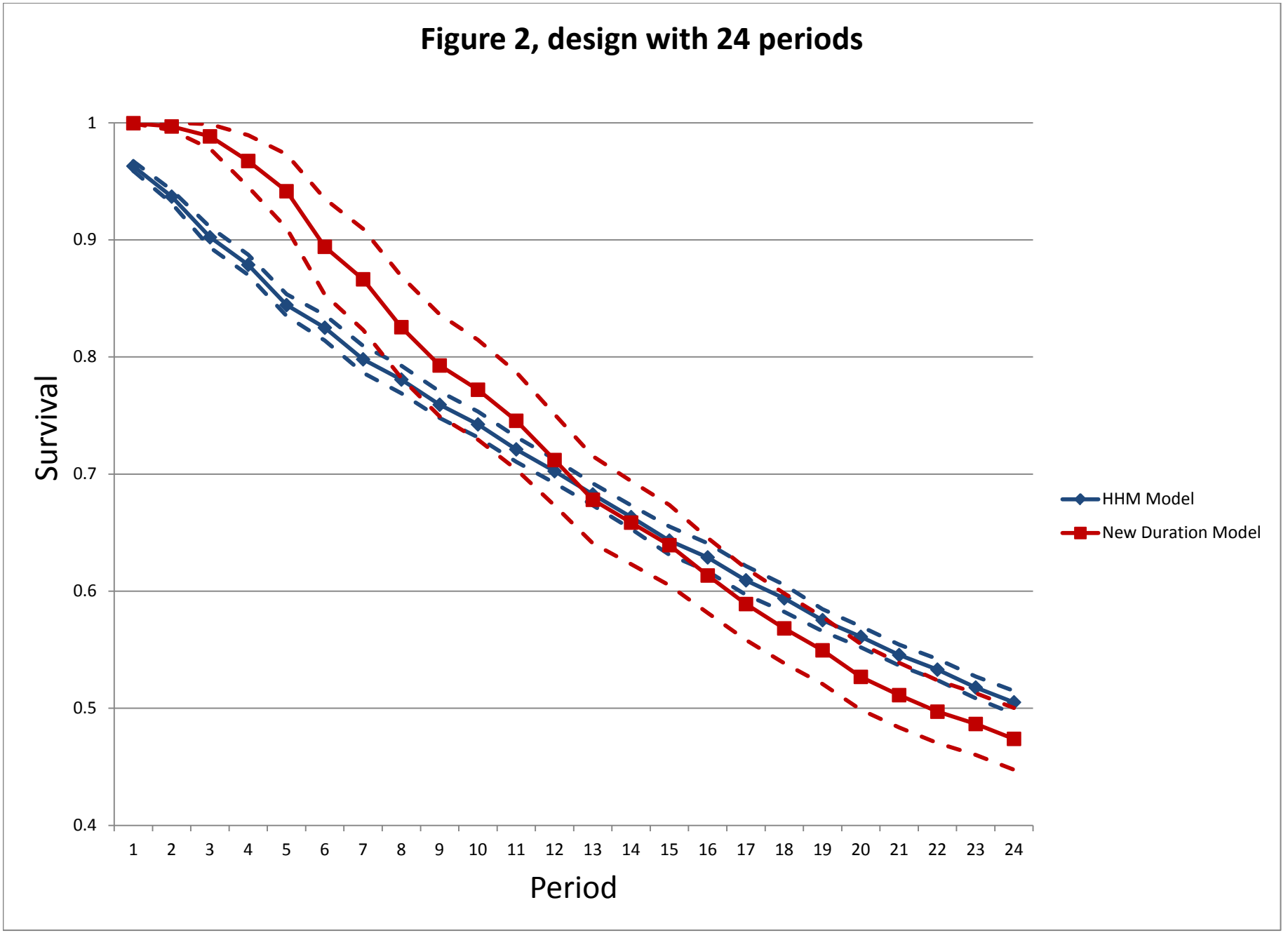
Figure 2, design with 24 periods
Survival
1
0.9
0.8
0.7
0.6
0.5
0.4
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Period
HHM Model
New Duration Model