Measurement Error Bias Reduction in Unemployment Durations

Montezuma Dumangane

THE INSTITUTE FOR FISCAL STUDIES
DEPARTMENT OF ECONOMICS, UCL
cemmap working paper CWP03/06

Measurement Error Bias Reduction in Unemployment Durations

Montezuma Dumangane*
CEMAPRE and ISEG-UTL

February 2006

Abstract. The impact of duration response measurement error is investigated by using small variance approximations. The inconsistency of GMM estimators that ignore measurement error is studied for both single spell models with right censoring, and for a two spell lagged duration dependence model. The results suggest a corrected GMM estimator for the error free and measurement error distributions. When the error free density is known, identification is achieved by using the moment condition that defines the measurement error sensitive specification score test. The results are applied to unemployment duration data from the BHPS.

Keywords: Measurement error, duration analysis, parameter approximations, GMM, score test, unemployment duration.

JEL Classification: C41, C51.

1. Introduction

Event histories are frequently constructed from recall data, and as a result the distribution of observed durations often differs from the distribution of the true durations due to contamination with measurement error. This leads to a distortion in the properties of duration distributions, for example typically having higher variance, differently shaped density and distorted hazard duration dependence.

Few attempts have been made to develop statistical procedures concerned with correcting for different types of duration response measurement error. Romeo (1997) uses a functional error-in-variables model and Bayesian techniques to estimate the true unobserved durations from multiple observations, which are used in a second stage as input to estimate the parameters of a Weibull model. Abrevaya and Hausman (1999) use the monotone rank estimator of Cavanagh and Sherman (1998) to produce consistent estimates of the covariate coefficients up-to-scale; this estimator does not require specification of a measurement error model but does not provide an estimator for parameters associated with duration dependence. Skinner and Humphreys (1999) derive an exact result for the Weibull model assuming a particular form for the measurement error and knowledge of its variance and study its bias, properties using small variance approximations.

*Department of Mathematics, ISEG, Technical University of Lisbon. Rua do Quelhas, 6, 1200-781 Lisboa, Portugal. E-mail: mdumangane@iseg.utl.pt, Tel:+351 21 3925876, Fax:+351 21 3922782.

This paper considers problems in which measurement error and error free durations are independently distributed. The main focus is to examine how the form of duration dependence of the hazard is affected by measurement error. Following the general approach of Chesher (1991), some general results concerning the impact of measurement error can be obtained leaving unspecified the distribution of measurement error by using small parameter asymptotic approximations. These are functions of the error free density and the variance of the measurement error only. They can be used to derive approximations for the effect of duration response measurement error on the probability limit of GMM-estimators constructed ignoring its presence. For particular parametric models, this measure gives important quantitative and qualitative information on the impact of measurement error on parameter estimates. The particular cases of the Log–logisitc, Weibull and a two-spell Exponential model with lagged-duration-dependence are studied. In the first two cases measurement error produces always attenuation bias on all parameter estimates, and its extent is shown to depend on the shape parameter and the proportion of censored observations. The third example shows that when measurement is correlated across spells, attenuation bias is just one of the possible outcomes. In a regression context this corresponds to a situation where dependent variable and covariate are both mismeasured and the errors are correlated. These results are extremely useful as they tell researchers in which situations measurement error is potentialy hazardous or when it can be "ignored".

The previous results suggest an approximate bias corrected GMM estimator of the parameters of the error free distribution, similar to Chesher (2000) for covariate measurement error. The general idea of this GMM-estimator is that if a model is characterized by a set of moment conditions that are not satisfied under certain misspecification (like measurement error), it is possible to find functions of the data that involve the unknown parameters that approximately correct the bias in the original moment conditions.

There is some related literature in the covariate measurement error problem concerning the correction of estimating equations. The conditional score method of Stefanski and Carroll (1987) assumes additive normally distributed measurement error and is applied to generalized linear models (see McCullagh and Nelder, 1989). The method of corrected score equations of Stefanski (1989) and Nakamura (1990) also assumes normality, and is extended in Buzas and Stefanski (1995) for certain generalized linear models. Bounaccorsi (1996) develops an estimator that unifies some of the previous approaches for a specific class of models.

When the density of the error free duration is assumed to be known, the estimation procedure proposed here does not require auxiliary data, often needed to identify the relevant parameters of the measurement error distribution. These are jointly estimated with the parameters of the error-free distribution by considering a GMM-estimator based on an extended score vector. This additional moment condition defines the measurement error specification score type test of Chesher, Dumangane and

Smith (2002) and Dumangane's (2000) extension for multiple-spells-single-destination (MSSD) models (see Chapter 3). An interesting consequence of this procedure is that score tests are rehabilitated from their major disavantage, namely that they give no constructive information on the structure of the model under the alternative.

This estimator is only approximately consistent, as the estimators in Chesher (1998, 1999), Wolter and Fuller (1982) and Carroll and Stefanski (1990) for the error-in-variables linear regression, the estimator in Chesher and Santos Silva (2002) for the multiple discrete logit model with uncontrolled taste variation, and Skinner and Humphreys (1999) estimator when the measurement error distribution is incorrectly specified.

As pointed out in Chesher, Dumangane and Smith (2002), when the true duration distribution belongs to the scale parameter family of distributions, multiplicative measurement error is statistically equivalent to scale parameter heterogeneity. As such, this estimator also allows for any unaccounted stochastic variation coming from the scale parameter. Therefore this estimator is an alternative to the parametric method proposed by Lancaster (1979) for the Weibull model that assumes a Gamma distributed random term in the scale parameter. It also belongs to the class of estimators for proportional hazards with unspecified unobserved heterogeneity that assumes a known parametric form for the baseline hazard, like those in Heckman and Singer (1984) and Honoré (1990). Horowitz (1999) extends these results by showing how to estimate non-parametricaly the baseline hazard function and the distribution of the unobserved heterogeneity. The importance of developing semiparametric estimators is outlined in Lancaster and Nickell (1980) and Heckman and Singer (1984). They alert to the possible missperception in the form of duration dependence induced by misspecification of the distribution of the random term.

As pointed out in Chesher (2001), since identification in this model requires a parametric assumption on the distribution of the error free duration, the procedure proposed here is presented mainly as a means of providing sensitivity analysis in the following sense: if the error free duration were as hypothesized and if there were measurement error, what would be the values of the parameters of the error free distribution and of the measurement error variance?[1]

The paper is organized as follows. Section 2 presents the assumptions of the measurement error model, recalls briefly the small parameter asymptotic approximations for single spell distributions derived in Chesher Dumangane and Smith (2002), and presents an extension for a favourable case of multiple-spell-single-destination model. The effect of measurement error on the hazard function is also described and illustrated. Section 3 derives the approximate probability limit of the inconsistent estimator and presents some examples. Section 4 derives the approximate bias cor-

[1] The issue of identification is not pursued in this study. In the related literature of neglected heterogeneity this issue is discussed in Lancaster (1979), Lancaster and Nickell (1980), Heckman and Singer (1984), Heckman (1991), and Elbers and Ridder (1982). Heckman and Taber (1994) list identification proofs for mixed proportionate hazard models.

rected GMM estimator. Section 5 presents some Monte Carlo on the performance of the estimator. Section 6 applies the estimator to mismeasured unemployment duration from the BHPS survey. Section 7 concludes.

2. The effect of measurement error

2.1. Single spell single destination. Let T be a scalar, non-negative-valued random variable, taken to represent the time to exit from a given state, with density function $f_T(\cdot)$ and survival function $\bar{F}_T(\cdot)$. These functions may depend upon a vector of observed covariates, X, but this dependence is not made explicit at present. Let the error-contaminated duration be $S = T \times V$ where $V \in [0, \infty)$ is a multiplicative measurement error continuously distributed independently of T with density function $f_V(v)$[2].

Under this conditions, Chesher, Dumangane and Smith (2002) demonstrate that the small parameter asymptotic approximations for the density and survival functions of S are[3]

$$f_S(s) \simeq f_T(s) + \frac{\sigma^2}{2}\left(f_T(s) + 3sf'_T(s) + s^2 f''_T(s)\right) \tag{1}$$

$$\bar{F}_S(s) \simeq \bar{F}_T(s) + \frac{\sigma^2}{2}\left(s\bar{F}'_T(s) + s^2\bar{F}''_T(s)\right). \tag{2}$$

where the second results from integration of (1). Here and later "$\simeq$" denotes an approximation error of order $o(\sigma^2)$ where $\lim_{\sigma \to 0} \frac{o(\sigma^2)}{\sigma^2} = 0$.

2.2. Multiple spell single destination models. Consider now a multiple-spell-single-destination process. A leading example is an individual that goes through a sequence of unemployment spells. The process is described by a sequence of calendar dates at which entry and exit to the states occurred. Let the sequence of R true durations in the state derived from those calendar dates be represented by the R-vector $\mathbf{T} = (T_1, T_2, ..., T_R)$. Assume the distribution of the error-free process has joint density function $f_{\mathbf{T}}(\mathbf{t})$, given by the product of the R conditional densities

$$f_{\mathbf{T}}(\mathbf{t}) = \prod_{j=1}^{R} f_{T_j|\mathbf{T}_{j-1}}(t_j|\mathbf{t}_{j-1}), \qquad f_{T_1|\mathbf{T}_0}(t_1|\mathbf{t}_0) = f_{T_1}(t_1) \tag{3}$$

Let $\mathbf{U} = (U_1, U_2, ..., U_R)$ be the measurement error vector distributed independently of $\mathbf{T}$, with joint continuous density $f_{\mathbf{U}}(\mathbf{U})$, satisfying $E(U_j) = 0$, $Var(U_j) = 1$ and $E(U_j U_l) = \rho_{jl}$, $j, l = 1, ..., R$. Let $\mathbf{S} = (S_1, S_2, ..., S_R)$ be the R-vector of error contaminated durations generated according to the measurement error model $\log S_j =$

[2]Since T is non-negative, multiplicative measurement error is the leading case of interest. The independence assumption is of course restrictive but also generates a leading case of interest.

[3]Here and later $'$, $''$ etc., indicate derivatives of functions in the sense that $f''_T(s) = \nabla_{tt} f_T(t)|_{t=s}$.

$\log T_j + \sigma_j U_j$ [4]. This error model is not valid when $\sum_{k=1}^{R} T_k = \sum_{k=1}^{R} S_k$, i.e., the age process is known or the spells are contigous as it imposes a specific form of correlation between the error terms and between these and the true sequence of durations[5]. The density of $\mathbf{S}$ is the $R-$folded integral

$$\int \cdots \int \prod_{j=1}^{R} f_{\mathbf{T}}(\mathbf{a})\, f_{\mathbf{U}}(\mathbf{u}) du_1 ... du_R, \tag{4}$$

where $\mathbf{a}$ is a R vector with elements $a_j = s_j \exp(-\sigma_j u_j)$. Let Σ be the $(R \times R)$ matrix with element σ_{kl} if $k \neq l$ and σ_k^2 if $k = l$. An approximation to the joint density of $\mathbf{S}$ can be deduced, by Taylor series expansion of (4) around $(\sigma_1, \sigma_2, ..., \sigma_R) = 0$, and collecting terms using the assumptions made on $\mathbf{U}$ we obtain,

$$\begin{aligned} f_{\mathbf{S}}(\mathbf{s}) \quad \simeq \quad & f_{\mathbf{T}}(s) + \frac{1}{2}\iota'\Sigma\iota f_{\mathbf{T}}(\mathbf{s}) + tr(\Sigma(\iota'\iota)\mathrm{diag}(\mathbf{s})\mathrm{diag}(F_{\mathbf{T}}^{(1)})) + \\ & + \frac{1}{2} tr(\Sigma(\mathrm{diag}(\mathbf{s})\mathrm{diag}(F_{\mathbf{T}}^{(1)})) + \frac{1}{2}\iota'\Sigma \otimes (\mathbf{s}'\mathbf{s}) \otimes F_{\mathbf{T}}^{(2)}\iota \end{aligned} \tag{5}$$

where $F_{\mathbf{T}}^{(1)} = \partial f_{\mathbf{T}}(\mathbf{t})/\partial \mathbf{t}$, $F_{\mathbf{T}}^{(2)} = \partial^2 f_{\mathbf{T}}(\mathbf{t})/\partial \mathbf{t} \partial \mathbf{t}'$, and ι is a $(R \times 1)$ vector of ones.

This expression is a generalization of (1) that accounts for correlated measurement error. Again it depends on the curvature properties of the error-free joint density function through its first and second partial derivatives.

2.3. The proportional representation. The measurement error model assumes that the observed duration is a random proportion of the true duration. As such the approximation to the survival function of S can be written as,

$$\bar{F}_S(s) \simeq \bar{F}_T(s)(1 + \tfrac{\sigma^2}{2}\mathrm{El}_{\bar{F}_T}(s)[1 + \mathrm{El}_{f_T}(s)]), \tag{6}$$

where $\mathrm{El}_{f_T}(t) = d\log f_T(t)/d\log t$ and $\mathrm{El}_{\bar{F}_T}(t) = d\log \bar{F}_T(t)/d\log t$ are respectively the elasticities of the density and the survival with respect to T. They measure the percentage response on those functions induced by a percentage change in T, and are responsible for, respectively, a scale and a sign effect of measurement error on the survival. Study of this expression allows to understand the distortion introduced by measurement error in a generic framework. Let T have a density function with at most one mode, then:

1. No matter the shape of the survival function, $\mathrm{El}_{\bar{F}_T}(t)$ is always negative and a decreasing function of T. It starts by being inelastic, that is $|\mathrm{El}_{\bar{F}_T}(t)| < 1$, and

[4] A more realistic version of the measurement error model would allow for heteroscedasticity in the measurement error variance by specifying $\sigma_{ij} = \sigma_j m(ij)$ where $m(ij)$ is a decreasing function of j, as the recall effort is bigger for earlier spells. The individual subscript i is needed as different stages of the process might have happen in different points in time demanding a different recall effort for each individual.

[5] Dumangane (2000) Chapter 2 illustrates this problem for the simple two spell case.

may become elastic, i.e. $|\mathrm{El}_{\bar{F}_T}(t)| > 1$ as T goes to infinity. As such, $\mathrm{El}_{\bar{F}_T}(t)$ scales the effect of measurement error on the survival, being the effect larger where and when the survival is elastic.

2. The term in square brackets is responsible for the sign of the effect of measurement error on the survival function. Let $T = t_f^*$ be the duration at which the density of T is unit elastic and negative, i.e., $\mathrm{El}_{f_T}(t_f^*) = -1$:
    (a) At t_f^* the effect of measurement error is null. This happens in the decreasing part of the density.
    (b) To the left of t_f^* the density is always negative-inelastic, that is $\mathrm{El}_{f_T}(s) > -1$, and if it has a mode it may be positive for durations closer to the origin. It follows that for $T < t_f^*$ the error contaminated survival is below the error free survival.
    (c) On the right tail of the distribution the density is negatively elastic, i.e. $\mathrm{El}_{f_T}(s) < -1$, so that the effect of measurement error is to raise the survival function above the error free survival.

Regardless the shape properties of the distribution of T, if the density has at most one mode measurement error lowers the survival for T smaller than t_f^* and raises it after that point. The extent of the distortion depends on the specific elasticity properties of the distribution of T.

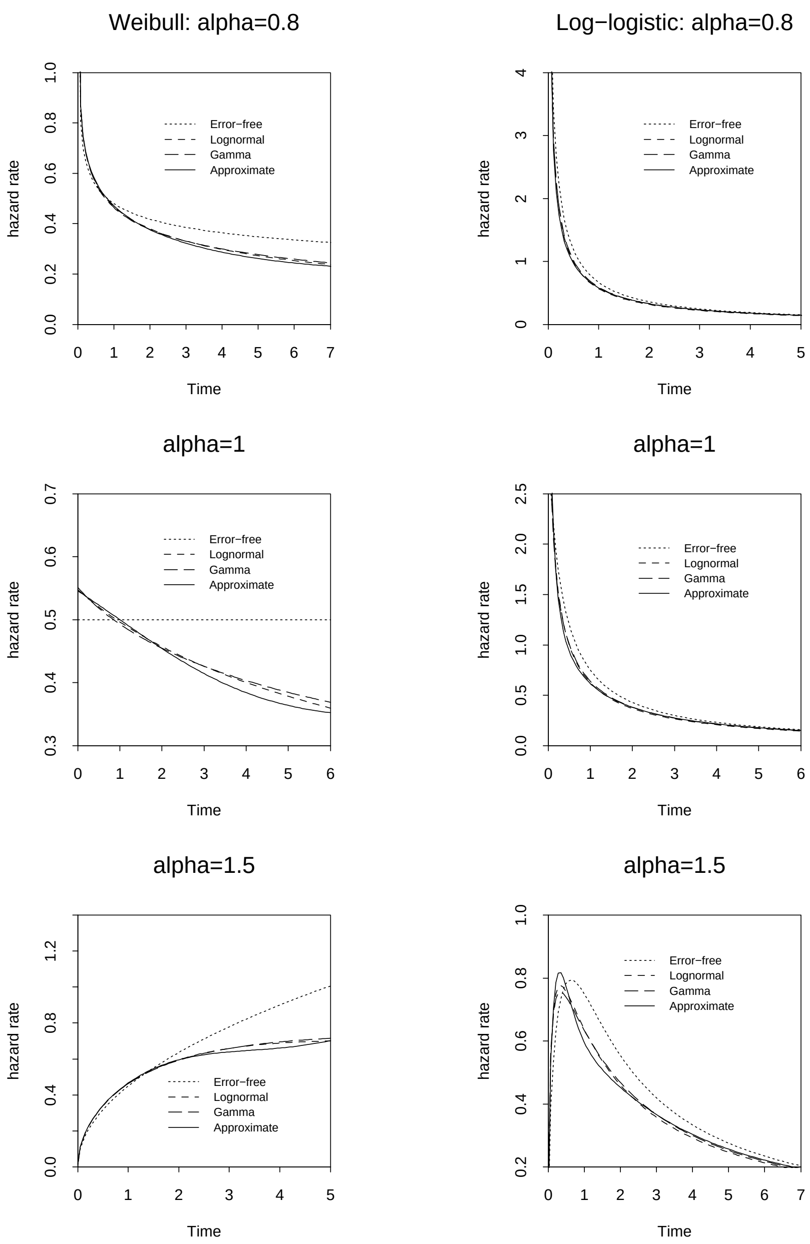
The impact of measurement error on the form of duration dependence of the hazard function follows directly. This is analysed in the next Section.

2.4. Approximate hazard functions. To study the effect of measurement error on the hazard function, $h_T(t)$, note that this function is equal to $-d\log \bar{F}_T(t)/dt$. It suffices to show how that transformation is affected by the shift in the survival function. Since the logarithmic function is monotonic and concave in its domain, implying that its derivative is a decreasing function and $\bar{F}_T(t)$ is itself a decreasing function of t it follows that

1. For short durations the effect of measurement error is to raise $h_S(s)$ above $h_T(s)$.
2. As $\bar{F}_T(s) > \bar{F}_S(s)$ for $s < t_f^*$, the error free survival must at this point be steeper than the contaminated survival, that is $d\log \bar{F}_S(t_f^*)/dt > d\log \bar{F}_T(t_f^*)/dt$. It then follows that $h_T(t_f^*) > h_S(t_f^*)$. A simple continuity argument shows that the two hazard functions must cross at some threshold $t_h^* < t_f^*$.

Figure 1 shows the impact of measurement error in a series of hazard specifications, and illustrate how informative the approximation can be on predicting the

Figure 1: Weibull and Log-logistic error-free, exact error-contaminated and approximate hazard functions, for $\alpha \in \{0.8, 1, 1.5\}$.



effects of measurement error on the form of duration dependence of the hazard function.

In all figures the error free hazard is either Weibull or Log-logistic with survival functions $\exp(-\lambda t^{\alpha})$ and $(1+\lambda t^{\alpha})^{-1}$ respectively. The multiplicative measurement error has a Lognormal or Gamma distribution with $E(\log V)=0$, and the parameters were chosen such that $Var(\log T)/Var(\log S)=0.8$. The range of T in all figures (horizontal axis) was chosen to satisfy $\bar{F}_T(t_{\max})=0.05$.

It is clear from the figures that the effect of measurement error changes the form of duration dependence in the hazard function in the way described by the approximation. In all plots the approximate hazard is a very good approximation to both error contaminated hazards for almost all s.

In both specifications the extent of the distortion depends positively on the shape parameter of the error free distribution. Moreover for the same value of α the Weibull hazards seem to be more sensitive to the measurement error.

The next Section investigates further this issues by giving results on the extent and nature of the inconsistency of GMM-estimators that ignore measurement error.

3. Approximate probability limit

Consider the class of single spell single destination models with right censored observations. Let t_i^* be the true lenght of time in the state for an individual. For a random draw from the population, if there was no measurement error the observed data would be $t_i=\min\{t_i^*,c_i\}$, $i=1,...,n$ where c_i is the censoring time for individual i . Let also $d_i=1(t_i^*<c_i)$ be the censoring indicator. If $d_i=1$, c_i is the potential censoring time (see Kalbfleisch and Prentice, 1980). Assume an independent random censoring (see Lawless 1982). Here and thereafter $E_T[\cdot|\phi=\phi]$ denotes expectations taken with respect to the error-free distribution at the parameter vector ϕ. Let the error free model be characterized by the set of moment conditions

$$E_T[g\left(T,\phi_0\right)|\phi=\phi_0]=0, \qquad g\left(t,\phi\right)=d\cdot g_1(t,\phi)+(1-d)\cdot g_0(t,\phi) \tag{7}$$

where $g_1(t,\phi)\equiv g(t,\phi|d=1)$ and $g_0(t,\phi)\equiv g(t,\phi|d=0)$ are $(q\times 1)$ vectors of functions, with $q\geqslant p$, depending on ϕ_0 the $(p\times 1)$ true parameter vector[6]. What follows is valid when all observations are uncensored, by letting c_i go to infinity.

3.1. Single spell models. Under the presence of measurement error, the observed data is $s_i=\min\{s_i^*,z_i\}$, $i=1,...,n$, where z_i is the error contaminated censoring time and d_i is assumed to remain unaffected by measurement error. Note that the distribution of the observed censoring times is non-informative about the parameter vector ϕ, this implies that the n-dimensional statistic $\{z_i\}_{i=1}^n$ is partially distribution constant for ϕ. By the partial conditionality principle (see Pace and

[6] In a parametric model those functions are respectively $g_1(t,\phi)=\nabla_{\phi}\log f_T(t,x,\phi)$ and $g_0(t,\phi)=\nabla_{\phi}\log\bar{F}_T(t,x,\phi)$.

Salvan, 1997), inference on the parameter vector $\theta = \{\phi, \sigma^2\}$ should treat the observed censoring times $Z_i = z_i$ as ancillary statistics on which inference should still be conditioned.

The density of S_i^* is $f_S(s, \theta_0)$ and θ_0 is the true parameter vector. Except when $g(t, \phi)$ are linear functions of $\log T$, measurement error changes the distribution of the data in such a way that the original moment conditions (7) are no longer satisfied[7].

The GMM estimator $\hat{\phi}_n$ that ignores the presence of measurement error is defined by

$$\arg\max_{\phi} \hat{Q}_n(\phi) = -\hat{g}_n(\phi)' \hat{W} \hat{g}_n(\phi) \tag{8}$$

where $\hat{g}_n(\phi) = n^{-1}\sum_{i=1}^n d_i g_1(s_i, \phi) + n^{-1}\sum_{i=1}^n (1-d_i) g_0(s_i, \phi)$, and $\hat{W}$ is a $(q \times q)$ positive semi-definite weighting matrix. By Lemma 2.3 in Newey and MacFadden (1994),the probability limit of $\hat{\phi}_n$, denoted by $\widetilde{\phi}(\theta_0)$, is the implicit solution of the $(q \times 1)$ system of equations

$$E_S[g(S, \widetilde{\phi}(\theta_0)|\theta = \theta_0] = 0 \tag{9}$$

If $\sigma_0^2 = 0$ this gives us (7) evaluated at ϕ_0. If $\sigma_0^2 \neq 0$, even if the measurement error distribution was specified, an explicit solution for $\widetilde{\phi}(\theta_0)$ is not trivial to find. Instead an approximation to $\widetilde{\phi}(\theta_0)$ can be constructed by first order Taylor series expansion around $\sigma_0^2 = 0$. First write (9) in the integral form

$$\int_0^z g_1(s, \widetilde{\phi}(\theta_0)) f_S(s, \theta_0) ds + g_0(z, \widetilde{\phi}(\theta_0)) \bar{F}_S(z, \theta_0) = 0 \tag{10}$$

Secondly, following the general approach of Kiefer and Skoog (1984) and upon replacing $f_S(s, \theta_0)$ and $\bar{F}_S(z, \theta_0)$ by its approximations, the term $\partial\widetilde{\phi}(\theta_0)/\partial\sigma^2$ at $\sigma^2 = 0$ in the expansion for $\widetilde{\phi}(\theta_0)$ is found by total differentiation of equation (10) with respect to σ_0^2 and $\widetilde{\phi}(\theta_0)$. Define $G_0 \equiv G(\phi_0)$ as the $(q \times p)$ matrix of expectations of the Hessian, i.e. $G_0 = E_T[\nabla_\phi g(S, \phi_0)|\phi = \phi_0]$, let $m_T(s, \theta)$ and $M_T(s, \theta)$, be respectively the $O(\sigma^2)$ terms in approximations (1) and (2); and let also $b(\theta_0) = b^a(\theta_0) + o(\sigma_0^2)$ where

$$b^a(\theta) = \int_0^z g_1(s, \phi) m_T(s, \theta) ds + g_0(z, \phi) M_T(z, \theta) \tag{11}$$

The $(q \times 1)$ vector function defined above is the approximate bias function in the moment conditions induced by measurement error which satisfies $b^a(\phi_0, 0) = 0$.

It follows that the GMM-estimator has probability limit given by

$$\widetilde{\phi}(\theta_0) \simeq \phi_0 - \left(G_0' W G_0\right)^{-1} G_0' W \; b^a(\theta_0) \tag{12}$$

Expression (12) shows that, up to the order here considered, the probability limit of $\hat{\phi}_n$ is a linear combination of the bias in the moment conditions induced

[7] The Exponential distribution is a case where multiplicative measurement error does not affect the moment condition.

by measurement error. The usual particular cases apply here, namely (i) when $q = p$, the matrix G_0 is square and expression (12) can be further simplified to yield, $\widetilde{\phi}(\theta_0) \simeq \phi_0 - G_0^{-1} b(\theta_0)$; (ii) if the model is parametric then $\widetilde{\phi}(\theta_0)$ is the approximate probability limit of the Maximum Likelihood Estimator (MLE), $G(\phi_0) = -E_T[\nabla_{\phi\phi} \log f_T(T, \phi_0)|\phi = \phi_0]$ is the Information Matrix, and $b^a(\theta_0)$ is the approximate bias of the score vector.

Under standard regularity conditions (see for example Newey and McFadden, 1994) the naive estimator has a well defined limiting distribution

$$\sqrt{n}(\hat{\phi}_n - \tilde{\phi}(\theta_0)) = N\left[0, (\widetilde{G}_\theta' W \widetilde{G}_\theta)^{-1} \widetilde{G}_\theta' W \ \tilde{\Omega}_\theta \ W' \widetilde{G}_\theta (\widetilde{G}_\theta' W \widetilde{G}_\theta)^{-1}\right] + o_p(1) \qquad (13)$$

where $\tilde{\Omega}_\theta = E_S[g(S, \tilde{\phi}(\theta_0)) g(S, \tilde{\phi}(\theta_0))'|\theta = \theta_0]$ is the asymptotic variance of the moment conditions evaluated at $\tilde{\phi}(\theta_0)$, $G_\theta(\phi) = E_S[\nabla_\phi g(S, \phi)|\theta = \theta_0]$, and $\widetilde{G}_\theta \equiv G_\theta(\tilde{\phi}(\theta_0))$.

3.2. Multiple spell. Consider now the class of MSSD models for complete observations only. Let the functions of $\mathbf{T}$ that define the moment conditions under the error-free model be $\mathbf{g}(\mathbf{t}, \boldsymbol{\phi})$, and define $\boldsymbol{\sigma}$ as the $((R + R(R-1)/2) \times 1)$ vector with the distinct elements of Σ. The approximate probability limit of the naive estimator $\tilde{\boldsymbol{\phi}}(\boldsymbol{\theta}_0)$, will now be of the form

$$\tilde{\boldsymbol{\phi}}(\boldsymbol{\theta}_0) = \boldsymbol{\phi}_0 + \sum_{k=1}^{R} \left.\frac{d\boldsymbol{\phi}}{d\sigma_k^2}\right|_{\boldsymbol{\sigma}=\mathbf{0}} \sigma_k^2 + \sum_{k=1}^{R-1} \sum_{l=k+1}^{R} \left.\frac{d\boldsymbol{\phi}}{d\sigma_{kl}}\right|_{\boldsymbol{\sigma}=\mathbf{0}} \sigma_{kl} + o(||\boldsymbol{\sigma}||) \qquad (14)$$

Let $\tilde{\boldsymbol{\phi}}(\boldsymbol{\theta}_0)$ be the parameter vector that solves the implicit set of equations now defined by $E_{\mathbf{S}}[\mathbf{g}(\mathbf{S}, \tilde{\boldsymbol{\phi}}(\boldsymbol{\theta_0}))|\boldsymbol{\theta} = \boldsymbol{\theta_0}] = 0$. Using the approximation to the multiple spell joint density the approximate probability limit of the inconsistent GMM estimator for this class of models is given by an expression similar to (12), where the approximate bias function is replaced by $b^a(\boldsymbol{\phi}_0, \boldsymbol{\sigma}) = \sum_{k=1}^{R} \sigma_k^2 b^k(\boldsymbol{\phi}_0) + \sum_{k=1}^{R-1} \sum_{l=k+1}^{R} \sigma_{kl} b^{kl}(\boldsymbol{\phi}_0)$. The terms in the summations are respectively given by

$$\begin{aligned} b^k(\boldsymbol{\phi}_0) &= \int_0^\infty \dots \int_0^\infty g(\mathbf{s}, \boldsymbol{\phi}_0) \nabla_{\sigma_k^2} f_{\mathbf{S}}^a(\mathbf{s}, \boldsymbol{\phi}_0) ds_R \dots ds_1 \qquad (15) \\ b^{kl}(\boldsymbol{\phi}_0) &= \int_0^\infty \dots \int_0^\infty g(\mathbf{s}, \boldsymbol{\phi}_0) \nabla_{\sigma_{kl}} f_{\mathbf{S}}^a(\mathbf{s}, \boldsymbol{\phi}_0) ds_R \dots ds_1 \end{aligned}$$

and $G(\phi_0) = \int_0^\infty \dots \int_0^\infty \nabla_\phi \mathbf{g}(\mathbf{s}, \boldsymbol{\phi}_0) f_{\mathbf{T}}(\mathbf{s}, \boldsymbol{\phi}_0) ds_R \dots ds_1$.

These results are now applied to some popular parametric models to see how informative the approximate probability limit of the MLE can be in describing the impact of measurement error.

Example 1. *: Flow-sample right-censored Weibull and Log-logistic hazard*

Consider the conditional Weibull and Log-logistic hazard functions,

$$h_T^W(t,x,\alpha,\beta) = \alpha\exp\{\beta' x\}t^{\alpha-1}, \quad \alpha>0 \qquad (16)$$

$$h_T^{LL}(t,x,\alpha,\beta) = \frac{\alpha\exp\{\beta' x\}t^{\alpha-1}}{1+\exp\{\beta' x\}t^{\alpha}}, \quad \alpha>0$$

Let in both cases the parameter β be partitioned in $\beta = (\beta_0\ \beta_1')$ and redefine β_0 so that x may be taken to have population mean zero and covariance matrix Σ_x.

Consider maximum likelihood estimation of the parameter vectors $\phi = \{\alpha, \beta\}$, allowing for the presence of independent right censoring. Except for the intercept in the Weibull model, in both specifications $\tilde{\phi}(\theta_0) \simeq k_j\phi_0$ with $j = W, LL$.

Figure 2 plots the approximate proportional bias k_j, against the conditional censoring proportion $\Pr(d = 0|c)$. Here c was made to vary to produce censoring proportions within the range of $[0, 0.8]$. The plots are entirely determined by the ratio $Var(\log T)/Var(\log S)$, and are invariant to β. Except when there is no censoring, direct application of expression (12) for the approximate probability limit does not have a closed form, and therefore numerical integration was needed.

It is clear that in both models duration response measurement error always dampnes the form of the duration dependence and attenuates the impact of covariates in the hazard function. For the Weibull model the inconsistency is a decreasing function of the censoring proportion, whereas in the Log-logistic the relation is non-monotonic.

The figure intercept corresponds to absence of censoring. In this case direct application of (12) yields expressions that are up to the order $o(\sigma^2)$ equivalent to

$$k_W = \frac{\psi'(1)}{\psi'(1)+\alpha_0^2\sigma_0^2} \qquad (17)$$

$$k_{LL} = \frac{1+2\psi'(1)}{1+2\psi'(1)+3\alpha_0^2\sigma_0^2/4}$$

where $\psi'(a)$ is the digamma function, and $\psi'(1) = \pi^2/6$. From (17) it is easy to see that the attenuation effect on the slope of the hazard function is determined by both σ^2 and α (the degree of log-convexity of the Weibull density). For the Weibull hazard the right hand side of (17) is just $Var(\log T)/Var(\log S)$. This is similar to the result in Lancaster (1990) for the approximate proportional bias of the MLE under the presence of proportionate hazard heterogeneity, with σ_H^2 -the variance of the random term- replaced by $\alpha^2\sigma^2$[8]. A similar resulta can also be found in Skinner and Humphreys (1999), since measurement error is implicitly treated there as neglected heterogeneity.

[8] As noted in Chesher, Dumangane and Smith (2002), if the distribution of T belongs to the scale parameter family of distributions, multiplicative measurement error is equivalent to scale parameter heterogeneity, and in the special case of the Weibull distribution it is also equivalent to proportionate hazard heterogeneity.

By comparing both expressions in (17), it can easily be seen that when there is no censoring the Log-logistic[9] specification is more robust than the Weibull, in the sense that the same relative amount of measurement error induces a smaller proportional bias in its parameter estimates, but as the proportion of censoring increases the opposite hapens.

In both cases the impact of measurement error is scaled by the form of duration dependence of the hazard function. These results are in agreement with the conclusions drawn in the previous section.

Example 2. *Two-spell Exponential lagged duration dependence*

Consider $R = 2$ and a lagged duration dependence model with Exponentially distributed spells, with scale parameters

$$\log \lambda_1 = \gamma_{01} + \gamma_{11}' x, \qquad \log \lambda_2 = \gamma_{02} + \gamma_{12}' x + \delta \log t_1 \tag{18}$$

The lagged duration coefficient is such that $Cov(\log T_1, \log T_2) = -\delta\psi'(1)$. If $\delta = 0$ this is an occurrence dependence model.

Assume that complete observations on $\{T_1, T_2\}$, from the flow of entrants in the first stage were used to compute maximum likelihood estimates of $\gamma_k = \{\gamma_{0k,}\gamma_{1k}'\}$, for $k = 1, 2$ and δ. Define $m_1 = \gamma_{02} - \psi(1)$ and $k_1 = \delta^2\sigma_1^2 + \sigma_2^2 - 2\delta\sigma_{12}$. The approximate probability limit of the MLE is,

$$\begin{pmatrix} \tilde{\gamma}_{01}(\boldsymbol{\theta}_0) \\ \tilde{\gamma}_{11}(\boldsymbol{\theta}_0) \\ \tilde{\gamma}_{02}(\boldsymbol{\theta}_0) \\ \tilde{\gamma}_{12}(\boldsymbol{\theta}_0) \\ \tilde{\delta}(\boldsymbol{\theta}_0) \end{pmatrix} \simeq \begin{pmatrix} \gamma_{01} - \frac{\sigma_1^2}{2} \\ \gamma_{11} \\ \gamma_{02} - k_1 - m_1 \frac{\delta_0\sigma_1^2 + \sigma_{12}}{\psi'(1)} \\ \gamma_{12} - \gamma_{11} \frac{\delta_0\sigma_1^2 + \sigma_{12}}{\psi'(1)} \\ \delta_0 - \frac{\delta_0\sigma_1^2 + \sigma_{12}}{\psi'(1)} \end{pmatrix} \tag{19}$$

The following points are of interest:

1. Since there is no form of duration dependence in the first spell the regressor coefficients are still consistently estimated[10].

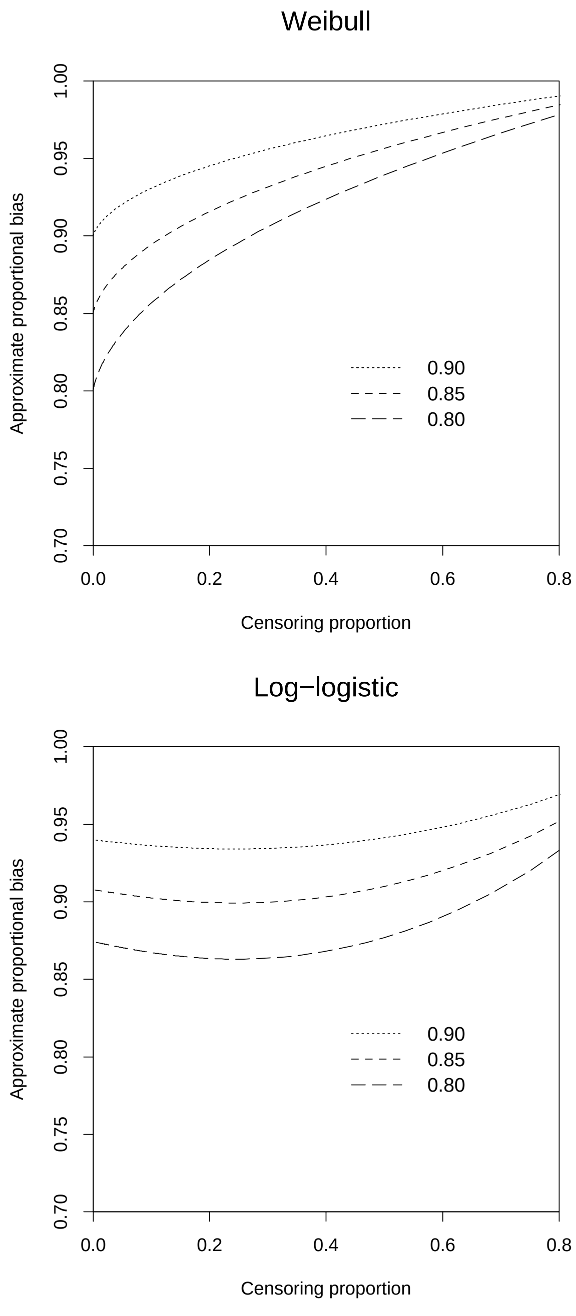
2. In this specification, the correlation between the measurement errors may lead to a missperception of the lagged duration coefficient sign. Consider the $o(||\boldsymbol{\sigma}||)$ equivalent expression for the approximate proportional bias of δ

$$\frac{\tilde{\delta}(\boldsymbol{\theta}_0)}{\delta_0} = \frac{\psi'(1)}{\psi'(1) + \sigma_1^2 + \sigma_{12}/\delta_0} \tag{20}$$

[9] For this model $Var(\log T) = 2\psi'(1)/\alpha^2$.

[10] This is because the score for this parameter is a linear function of $\log T_1$.

Figure 2: Approximate proportional inconsistency of α and β_1, as a function of the censoring proportion for $Var(\log T)/Var(\log S) \in \{0.80, 0.85, 0.90\}$.



Whenever the covariance between the log durations has the same sign of the covariance between the measurement errors, the result will be an attenuation effect, otherwise the inconsistency may lead to a sign change.

3. In the simple case of $\delta = 0$, estimated duration dependence might be the consequence of correlated measurement error and therefore totally spurious.
4. Only the coefficients associated with covariates that appear in the first spell are affected by measurement error.
5. The extent of the inconsistency in the covariate coefficient is determined by the extent of the inconsistency in the lagged duration coefficient weighted by the covariate coefficient in the first spell.
6. If the same set of covariates affect the two duration distributions in the same fashion, then the proportionate bias will be as before, equal to the proportionate bias in the lagged duration coefficient.
7. All that was said about the misperception of lagged duration dependence applies to the covariate coefficients with the additional complication introduced by the coefficient γ_{11}. The potential misperception of the sign of the second spell coefficients is a possible consequence of measurement error.

The result from these first two sections show how GMM estimators are inconsistent when the dependent variable is contaminated with measurement error. The inconsistency arises because the moment conditions that define the error free model are not satisfied under the contamination. The next section uses this result to derive an approximate bias corrected GMM estimator.

4. Bias corrected GMM estimator

4.1. Single spell models. Let the model for T be characterized by the set of moment conditions defined in (7). The estimator proposed here is based on the principle that the moment conditions can be approximately corrected by functions of the observed data, and then used to construct a GMM estimator.

From the previous section it follows that $E_S[g(S,\phi) - b^a(\theta)|\theta = \theta] \simeq 0$. Write this moment condition as

$$\int_0^z (g_1(s,\phi) - b_1^a(s,\theta))f_S(s,\theta)ds - r_1(z,\theta) + (g_0(z,\phi) - b_0^a(z,\theta))\bar{F}_S(s,\theta) \simeq 0 \quad (21)$$

Because of the order of the approximation considered here, terms of order $O(\sigma^m)$ with $m > 2$ can be omitted. It follows from (11) that

$$b_0^a(z,\theta) = g_0(z,\phi)M_T(z,\theta)\bar{F}_T^{-1}(s,\phi)$$

As for $b_1^a(s,\theta)$ and $r_1(z,\theta)$, they solve the equation that equals the integral in (11) to

$$\int_0^z b_1^a(s,\theta) f_T(s,\phi) ds + r_1(z,\theta) \tag{22}$$

In the appendix it is shown that under independent random censoring the following expressions hold:

$$\begin{aligned} b_1^a(s,\theta) &= \tfrac{\sigma^2}{2}[s g_1'(s,\phi) + s^2 g_1''(s,\phi)] \\ r_1(z,\theta) &= \tfrac{\sigma^2}{2}\{[z g_1(z,\phi) - z^2 g_1'(z,\phi)] f_T(z,\phi) + z^2 g_1(z,\phi) f_T'(z,\phi)\} \end{aligned} \tag{23}$$

Then, the approximate structural bias function, $b^a(s,z,\theta)$, i.e. the vector function of the data that corrects the bias in the moment conditions, is given by

$$b^a(s,z,\theta) = d\ b_1^a(s,\theta) + (1-d)\ b_0^a(z,\theta) + r_1(z,\theta) \tag{24}$$

Under standard tail conditions for the density of T, the function $r_1(z,\theta)$ vanishes as $z \to \infty$, leading to the result for complete spell models. In this case knowledge of the distribution of T is not needed to correct the moment conditions.

It follows from (24) that the bias corrected moment conditions are

$$E_S[g^c(S,z,\theta)|\theta=\theta] \simeq 0, \qquad g^c(s,z,\theta) = g(s,z,\phi) - b^a(s,z,\theta) \tag{25}$$

where $g^c(s,z,\theta) \simeq g_S(s,z,\theta)$, and $E_S[g_S(S,z,\theta)|\theta=\theta]=0$, defines the exact unbiased moment conditions for the error contaminated model.

4.2. Multiple spell models. Consider now the multiple spell model in Section 2. In the appendix the structural bias function is shown to be

$$\mathbf{b}^a(\mathbf{s},\theta) = \tfrac{1}{2} tr(\Sigma \text{diag}(\mathbf{s}) \text{diag}(\mathbf{G}_\mathbf{T}^{(1)})) + \tfrac{1}{2}\iota'\Sigma \otimes (\mathbf{s}'\mathbf{s}) \otimes \mathbf{G}_\mathbf{T}^{(2)}\iota \tag{26}$$

where $\mathbf{G}_\mathbf{T}^{(1)} = \partial \mathbf{g}(\mathbf{t})/\partial \mathbf{t}$ and $\mathbf{G}_\mathbf{T}^{(2)} = \partial^2 \mathbf{g}(\mathbf{t})/\partial \mathbf{t} \partial \mathbf{t}'$. When $R=1$ and there is no censoring this leads to (24).

4.3. Identification and estimation. The measurement error bias corrected estimator can now be defined given a conditional density for T, $f_T(t,\phi)$, and a sample of i.i.d. observations on $\{s_i, z_i, d_i\}_{i=1}^n$. Let $g_1(t,\phi) = \nabla_\phi \log f_T(t,\phi)$ and $g_0(t,\phi) = \nabla_\phi \log \bar{F}_T(t,\phi)$.

If σ^2 is unknown, an additional moment condition is necessary to identify θ. Consider $D_{\sigma^2}(t,c,\phi)$, the score vector for the variance of the measurement error at $\sigma^2=0$, which satisfies $E_T[D_{\sigma^2}(T,c,\phi_0)|\phi=\phi_0]=0$. Let $D_{1,\sigma^2}(t,\phi)$ and $D_{0,\sigma^2}(t,\phi)$ denote its contributions for, respectively, complete and censored observations, derived from (1). In Chesher, Dumangane and Smith (2002) this moment condition was the basis to construct a measurement error specification test for $H_0: \sigma^2=0$. Define now the $(q+1 \times 1)$ extended score vector $g_e(t,c,\phi)' = (g_e(t,c,\phi)'\ D_{\sigma^2}(t,c,\phi))$.

The bias corrected GMM estimator proposed here is based on the $(q+1)$ set of moment conditions

$$E_S[g_e^c(S,z,\theta_0)|\theta=\theta_0] \simeq 0 \tag{27}$$

where $g_e^c(s,z,\theta)$ is the bias corrected extended score vector. Let the sample counterparts of the moment conditions be $\hat{g}_{e,n}^c(\theta) = n^{-1}\sum_{i=1}^n g_e^c(s_i,z_i,\theta)$. Under suitable regularity conditions that ensure existence and uniqueness, the proposed GMM estimator $\hat{\theta}_n^c$, is defined as $\arg\max \hat{Q}_{e,n}^c(\theta) = -\hat{g}_{e,n}^c(\theta)'\hat{g}_{e,n}^c$.

Under standard regularity conditions, (see for example Newey and McFadden 1994) the first order asymptotic distribution of the bias corrected GMM estimator is

$$\sqrt{n}(\hat{\theta}_n^c - \theta^a) = N\left[0,(G_e^a)^{-1}\ \Omega_e^a\ (G_e^a)^{-1\prime}\right] + o_p(1) \tag{28}$$

where $\theta^a = P\lim \hat{\theta}_n^c$, $\Omega_e^a = E_S[g_e^c(S,\theta^a)g_e^c(S,\theta^a)'|\theta=\theta_0]$ is the asymptotic covariance matrix of the approximate bias corrected extended moment conditions evaluated at θ^a, and $G_e^a = E_S[\nabla_\theta g_e^c(S,\theta^a)|\theta=\theta_0]$.

Since the estimator is derived from an objective function that omits terms of order $o(\sigma^2)$ in the moment conditions, unless $\sigma^2=0$, θ^a is not in general equal to θ_0, but as is shown in the appendix, $\theta^a = \theta_0 + O(\sigma^3)$, so that $\hat{\theta}_n^c$ has a smaller asymptotic bias than the inconsistent GMM-Estimator.

Figures (3) and (4) show the exact expectation of the uncorrected and approximate bias corrected extended score vector at the true parameter values as a function of the proportion of variance in the log duration due to measurement error. Two measurement error distributions were used, the Lognormal and the two parameter Gamma, to contaminate the Weibull and Log-logistic distributions. In all left panels there is 20% of censoring and in all right panels 50%. Except for the scale parameter with 50% censoring, the lines closer to the horizontal line always correspond to the bias corrected scores. In the plot for the shape parameter the bias in the uncorrected scores is an increasing function of the shape parameter. Despite that the quality of the correction is independent of α.

These plots suggest that, whenever the model is identified, the proposed estimator will be a considerable improvement on the maximum likelihood estimator that ignore measurement error.

Figure 3: Exact expectation of Weibull scores and approximate bias corrected scores with Lognormal (dotted) and Gamma (dashed) measurement error for $\alpha \in \{0.8, 1, 1.5\}$ and 20% and 50% censoring.

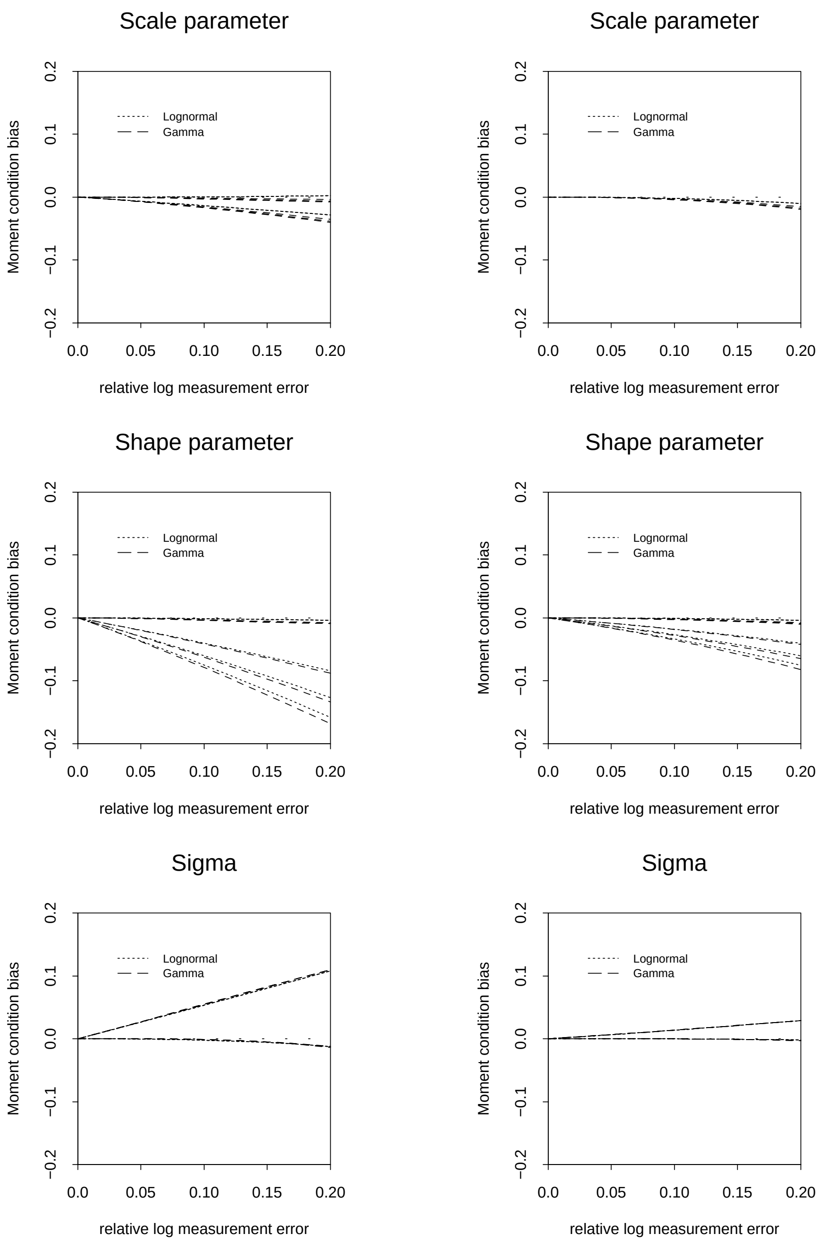
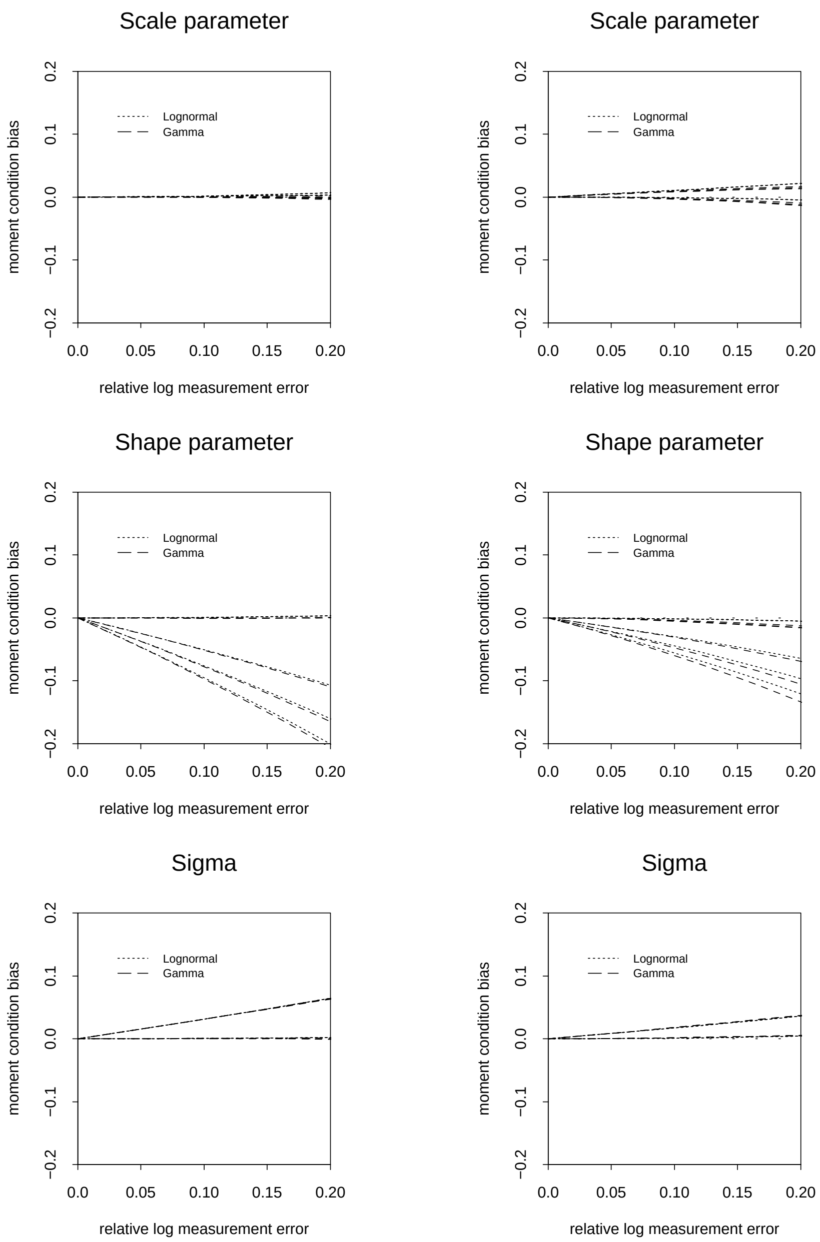


Figure 4: Exact expectation of Log-logistic scores and approximate bias corrected scores with Lognormal (dotted) and Gamma (dashed) measurement error for $\alpha \in \{0.8, 1, 1.5\}$ and 20% and 50% censoring.



4.4. Heteroskedastic measurement error. Being a memory, problem it is reasonable to assume that the distribution of measurement error should depend on some measure of "recall effort" that may differ across individuals. A simple and intuitive way of incorporating this idea is to specify a measurement error variance function that depends on the recall effort. Let the W be such a measure, observable and independent of T. Then $\sigma_i^2 = m(w_i, \pi)$ for some positive valued function $m(.)$.

Because W is assumed to be independent of T, the results on sections 2 and 3 are still valid with σ^2 replaced by $m(w, \pi)$. Two approaches are suggested for estimation.

Known variance function. The first approach requires the specification of the variance function. Let $m(w_i, \pi) = m(\pi_0 + \pi_1' w_i)$ be a positive valued differentiable function with, $m(0) = \sigma^2$ and finite $m'(0)$. A natural candidate for $m(.)$ is the exponential function. In any case if $b^a_{\boldsymbol{\sigma}}(s, z, \theta)$ denotes the approximate structural bias function associated with $D_{\sigma^2}(t, c, \phi)$, then estimation of $\pi' = (\pi_0, \pi_1')$ requires the additional estimating equations:

$$\begin{cases} n^{-1} \sum_{i=1}^n D^c_{\sigma^2}(s_i, z_i, \theta) = 0 \\ n^{-1} \sum_{i=1}^n D^c_{\sigma^2}(s_i, z_i, \theta) w_i' = 0 \end{cases} \tag{29}$$

where $D^c_{\sigma^2}(s, z, \theta) = D_{\sigma^2}(s, z, \phi) - b^a_{\boldsymbol{\sigma}}(s, z, \theta)$. As usual consistency requires correct specification of $m(.)$. The second approach tries to correct for this shortcome.

Unknown variance function. In this approach all that is required is that $m(w)$ be a monotonic function of w. Consider the thresholds for values of the recall effort variable $\{w_0, w_1, ..., w_p\}$ where the lower and upper limit may be infinity. Let $d_{ji} = 1(w_{j-1i} < w_i < w_{ji})$, $j = 1, ..., p$; then a semiparametric specification for then variance function is $\sigma_i^2 = \sum_{j=1}^p \sigma_j^2 d_{ji}$. The p additional estimating equations will now be $n^{-1} \sum_{i=1}^n D^c_{\sigma^2}(s_i, \theta) d_{ji} = 0$, $j = 1, ..., p$. Of course, results may be sensitive to the specification of the intervals but still independent from parametric assumptions.

5. MONTE CARLO RESULTS

This section reports Monte Carlo experiments designed to investigate the performance of the bias corrected GMM estimator. Models for flow sample data allowing the presence of random censoring (right censoring) will be considered. In these set of experiments the error free duration T^* has a two-parameter Weibull distribution or a two-parameter Log-logistic distribution, with conditional hazard functions given in (16). Since the correction on the moment conditions is independent of α, in all experiments $\alpha = 1$. The regressor is chosen to be $x = [1, x_1]$, with x_1 taking 50 values evenly spaced in $[-1, 1]$, and $\beta_0 = \beta_1 = 1$[11].

Allowing for right censored observation means that the true duration is $T_i = \min\{T_i, C_i\}$ with $d_i = I(T_i < C_i)$, where the censoring times C_i are i.i.d. random variables with survival function $\bar{H}(c_i|\lambda_i)$, independently distributed from T_i^*. Here

[11]Other fixed regressor designs were used and did not change the properties of the estimator.

$\bar{H}(c_i|\lambda_i) = \bar{F}_T(c_i, \alpha, \omega\lambda_i)$, for $\lambda_i = \exp(\beta' x_i)$, where ω is a constant chosen such that the proportion of censored observations δ, satisfies (on average)

$$\Pr(d = 0|C) = \Pr(T > C|\alpha, \lambda, \omega) = \delta \tag{30}$$

for $\delta \in \{0.2, 0.5\}$[12].

Since random variable C is only observed for censored observations, for complete durations the potential censoring times C_p must be defined, satisfying $C_p \geq T$. In these experiments $C_p = T$.

Since measurement error is only observed for complete observations, in this experiments the measurement error vector will be given by $V = dV_1 + (1 - d)V_0$, and $V_0 = \sqrt{k}V_1$, for $k = 0.8$[13].

The error-contaminated duration is $S = TV$ with $E[\log V] = 0$ where V is generated independently of T. In half of the experiments V is generated as Lognormal(μ, σ^2), in the other half V is generated as Gamma(m, n). The parameters μ, σ^2, m and n were chosen such that the logarithmic signal to noise ratio, $\rho = Var(\log T)/Var(\log S)$ belongs to $\{0.80, 0.90\}$. Each experiment employed 2,000 replications and sample sizes of 200 and 500 observations were used in all experiments.

Two estimators will be considered. The first is just the MLE that ignores measurement error. The second is the bias corrected GMM that estimates all unknown parameters.

5.1. No measurement error.

(Table 1 about here)

Table 1 reports the GMM estimates when there is no measurement error. The Weibull estimates are always mean and median unbiased, except for β_0 with 50% censoring where the mean bias is 2.9%. The Log-logistic estimates are always median unbiased but exhibit a slight mean bias especially for higher censoring. Nevertheless it performs fairly well.

---

[12] The corresponding values of ω were found by solving the equation

$$\omega : \qquad \Pr(Y > 1|\alpha, \lambda, \omega) = \delta$$

where $Y = T/C$, is a positive random variable with density function

$$f_Y^W(y) = \frac{\alpha\omega y^{\alpha-1}}{(\omega + y^\alpha)^2},$$
$$f_Y^{LL}(y) = \frac{\alpha k y^{\alpha-1}(2(\omega - y^\alpha) + (\omega + y^\alpha)(\alpha \log y - \log \omega))}{(y^\alpha - \omega)^3}$$

for respectively the Weibull and the Log-logistic distribution. Note that the distribution of Y is independent of λ.

[13] As $Var(V_0)/Var(V_1) = k$, this is a measure of the proportion of measurement error attributed to misreporting the starting data of the spell.

5.2. Under measurement error.

(Tables 2 and 3 about here)

Tables 2 and 3 report the maximum likelihood estimates under two different amounts of measurement error. They allow us to verify the accuracy of the predictions of the approximate probability limit. As expected measurement error produces an attenuation bias in all parameters. The proportional bias in α and β_1 are equal in the Weibull model, and also equal to the proportional bias in β_0 for the Log-logistic when measurement error is small. As predicted, the proportional bias is a decreasing function of the proportion of censoring for the Weibull, but increasing (in this range) for the Log-logistic. Although for $\rho = 0.9$ the proportional bias is small, for $\rho = 0.8$ in both models the proportional bias is considerable.

(Table 4 about here)

Table 4 show the bias corrected GMM estimator for the Weibull model. For $\rho = 0.9$ the estimates of β_0, β_1 and α are very close to one, deviating at most 3.1%. For these experiments the true variance of the log measurement error is 0.183. The GMM estimates are clearly downward biased and very imprecise.

For $\rho = 0.8$ the GMM estimates are clearly downward biased both for mean and median. Despite this behaviour they are still a clear improvement on the maximum likelihood estimates. For larger sample sizes, the bias of α reduces from 16.5% to 6.1% with 20% censoring, and from 15% to 8% with 50% censoring, but the accuracy loss with respect to maximum likelihood is very large. The variance of the log measurement error is now 0.41 and the GMM estimates recover up to 85% of it, but are severly median biased.

(Table 5 about here)

Table 5 show the results for the Log-logistic model.

When $\rho = 0.9$, at both sample sizes the means and medians of all estimates are very close to one. When $\rho = 0.8$ all estimates are slightly mean and median downward biased. For larger sample sizes the bias of α reduces from 12% to 3% with 20% censoring, and from 14% to 5% with 50% censoring. On average the standard deviations are two times the maximum likelihood standard deviations. The measurement error variances are 0.365 and 0.822 for respectively $\rho = 0.9$ and $\rho = 0.8$. In both cases the GMM estimate are (in the best cases) just slightly median biased. It recovers 91% of the true variance when $\rho = 0.9$ and $\delta = 0.20$, and 95% when $\rho = 0.9$ and $\delta = 0.50$. Despite that this estimates are still very imprecise.

These simulations suggest that the bias corrected GMM estimator is always an improvement relatively to the naive maximum likelihood. Its performance is clearly better under the Log-logistic distribution, being the increased variability an expected cost. Here the mean and median properties of the estimator are less affected by increasing quantities of measurement error. Given the moment condition bias showed in the previous section, the results for the Weibull model suggest that, as measurement error increases, parametric identification is not enough to identify the true parameter vector. In this sense for this specification this can only be a small variance procedure.

In the next section the theory developed here is applied in estimating the conditional distribution of unemployment durations.

6. Error contaminated unemployment durations

Several studies using the British Household Panel Survey (BHPS) have examined recall error in unemployment durations. Paull (1997, 2000) uses it to investigate the impact of recall error on labour market behaviour; Elias (1996) and Dex and McCulloch (1997) look at recall of unemployment in the BHPS employment-status history, and the Family and Working Lives Survey, and in the Labour Force Survey. Brendan (1997) studies the BHPS in the period between September 1990 and September 1991, for which both the wave 1 and wave 2 retrospective surveys provide information on unemployment, and found disagreements in the two surveys.

In this application data from the BHPS collected at wave 1 is used. The sample includes all male individuals that reported having experienced unemployment between 9/90 and the date of interview at first wave, which spanned till 12/91. For each individual, information on the start and exit dates of the reported unemployment spell is collected. For those who experienced multiple spells of unemployment in that reference period only the latest (closer to the interview date) is considered. Wave 1 also reports information about the individual characteristics, including income variables.

6.1. The model. The economic specification follows a reduced form approach, which implies the estimation of the parameters of the hazard function for the time to leaving unemployment, conditional on a set of key exogenous variables, such as unemployment benefits and other variables. Early examples of this approach are Lancaster (1979) and Nickell (1979a,b). Narendranathan, Nickell and Stern (1985) provide an excellent discussion on the effect of unemployment benefits in unemployment duration, referring to the work of Atkinson and Flemming (1978) and Atkinson et al. (1982).

The conditional distribution of time to leaving unemployment will be a function of the following variables:

1. Age: the logarithm of age.
2. Higher Qualification and Lower Qualification: Educational dummies identifying respectively, higher degree, first degree, teaching qualification and other higher qualification, from, CSE , commercial, GCE and nursing qualifications, apprenticeship and other lower qualifications. Both are zero for no qualifications.

3. Married: A dummy variable which equals one if the individual is married or lives as a couple.

4. Children: The number of dependent children in the household.

5. Local Unemployment Rate: The unmeployment rate at the metropolitan area of residence.

6. Income in Unemployment: The log of weekly benefits received by the individual from all sources -Unemployment and Supplementary Benefits, Family Income Support, Child Benefit and other government transfers- while unemployed, at time of exit from unemployment [14].

7. Income in work : The log of weekly estimated earnings specified as a function of work experience measures and other individual characteristics[15]. This involved estimation of a participation equation as described in the appendix to account for selection bias. This variable measures the mean of the wage offer distribution that faces the individual, and is interacted with an indicator variable that distinguishes whether an individual received any benefits while unemployed.

(Table 6 about here)

Table 6 shows some descriptive statistics of the variables used in this study before being transformed. As expected, those who never had any form of unemployment benefit are on average younger, more educated and experience smaller spells of unemployment.

This sample considers individuals that experienced unemployment between 9/90 and the date of interview, that extended until 31/9, the reference period for the wave 1 survey. As such, individuals from two distinct populations were sampled: those belonging to the stock of the unemployed at the calendar time T_0, the start of the reference period, and those who flow into unemployment after T_0. For both samples from these populations a complete or censored duration is recorded at time T_I, the date of interview for the i-ith individual. Such duration is obtained from information on the entry and exit dates from unemployment, T_E and T_X respectively, both collected retrospectively.

[14] In practice this should be a time varying covariate as the level of benefits vary during the unemployment spell, replacing it by a single value is a rough approximation.

[15] Traditional search theory postulates that this variable should have a positive effect on the probability of leaving unemployment. However, if the rate of job offers is a function of the mean wage, such that it is higher in segments of the labour market for which the mean wage is lower, than this variable could have a negative effect on the hazard rate. It may also capture the fact that high profile jobs have a greater competition for, therefore being more difficult for individuals in this cohort to exit unemployment.

In Dumangane (2000) the likelihood of the resulting sample of 510 male individuals (of which 60% were still unemployed at the date of interview) is shown to be[16],

$$f_T^*(t,x,\phi) = \frac{tf_T(t,x,\phi) + E[\triangle T]f_T(t,x,\phi)}{E[T] + E[\triangle T]} \tag{31}$$

Expression (31) gives the individual contribution for the likelihood of the sample as a weighted average of the likelihood of a stock sample and a flow sample. The size of $E[T]$ (the unemployment rate) relatively to $E[\triangle T]$ (the average length of the reference period for this survey) determines the weight assigned to each population contribution to the data. If the unemployment rate is high, than the sample scheme will be closer to a stock sample. A simplifying assumption on the form of the density (31) will be made, namely $f_T^*(t,x,\phi)$ will be assumed to belong to a known parametric family.

The censoring rule for each individual is such that the study ends at the date of interview T_{Ii}, which is independent across individuals, and independent of the spell length. As such the censoring and potential censoring times are $Z = T_I - T_E$.

Two alternative parametric hazard specifications will be considered, the two-parameter Weibull and the Log-logistic. The first can be used to test whether unemployment duration is a time dependent process, and is expected to produce a decreasing hazard rate. The second allows for non-monotonic hazard functions, representig an unemployment process in which initially the risk of leaving unemployment increases, then reaches a peak after which unemployment becomes persistent.

6.2. Maximum likelihood estimates. Being a retrospective survey, measurement error is an issue to take into account. In this cohort, 25% of the population started the spell of unemployment in the six months before the data of interview, 37.5% between 6 months and one year, and the remaining 37.5% more than one year before the date of interview[17].

(Table 7 about here)

Table 7 reports maximum likelihood and GMM estimates for both specifications[18]. The MLE estimate of the Weibull shape parameter suggests a decreasing hazard for leaving unemployment and the constant hazard rate model is rejected. This result is consistent with the search theory postulate, that asserts that as individuals spend time in unemployment the rate of job offers decreases such that the progressive

[16] I am in debt with Andrew Chesher for helping deriving this sampling distribution.

[17] In fact some of this individuals could not recall accurately the dates at which unemployment occurred, and for those only the month and year is recorded, being the date considered the 15th of the reported month.

[18] Bootstrap standard errors, taking into account that one of the regressors is estimated, were computed using 1000 bootstrap replications. The null hypothesis for α is $H_0 : \alpha = 1$. As the hypothesis $H_0 : \sigma^2 = 0$ lies on the boundary of the parameter space, the test is based on t^2 and the asymptotic 5% critical level $c^*_{0.05}$ solves, $\Pr(\chi^2(1) < c_{0.45}) + 0.5 = 0.95$ (see Godfrey, 1988).

expected reduction in the reservation wage fails to compensate for a falling rate of job offers. The Log-logistic MLE estimate shows that unemployment duration is described by a non-monotonic hazard, being the hypothesis of a monotonic decreasing hazard rate rejected.

Under the null hypothesis of homogeneity, the integrated hazard vector are n realizations of a mean-one Exponential variate. This allows to perform residual analyses (see Lancaster and Chesher, 1985) as a means to study the quality of the fit in both parametric specifications.

Figure 5 shows the plots for the maximum likelihood estimates of both specifications. The plot for the Log-logistic specification falls everywhere closer to the 45 degree line, suggesting that this model is a better approximation to this data than the Weibull specification.

In the bottom of Table (7) is reported the test statistic for the efficient version of the measurement error specification test for the null hypothesis $H_0 : \sigma^2 = 0$. The calculation of the efficient version of the test involved numeric computation of the asymptotic variance of the test[19]. At a 0.05 nominal level, the null hypothesis is clearly rejected in the Weibull model, suggesting the presence of measurement error, but it is not rejected for the Log-logistic model at the same nominal level.

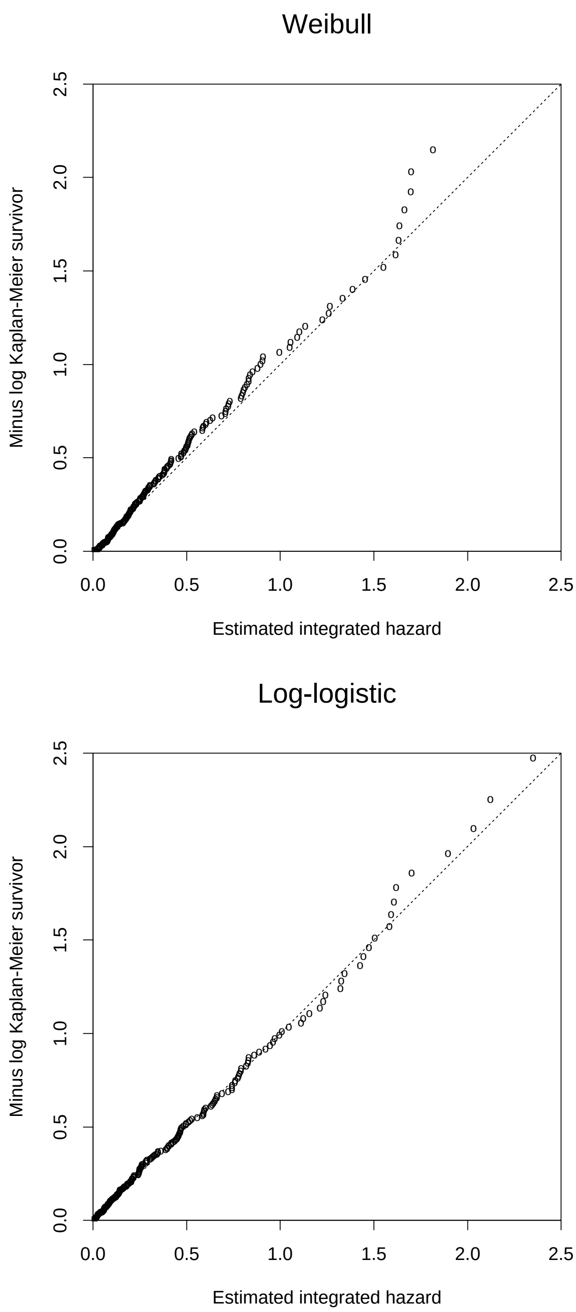
6.3. The bias corrected estimates. The Weibull bias corrected GMM estimates indicate that the measurement error variance is not significantly different from zero. Despite that, the coefficients (and in particular α) are now bigger and the exponential model is not rejected at the usual signifcance levels. This results are not consistent and may reflect a misspecification of the error free distribution.

The bias corrected Log-logistic estimate of α is consistent with the effect of measurement error in this parameter. In fact,

1. The correction has the right sign, and according to the estimate of the variance of measurement error 15% of the observed variation in the log durations is attributed to this misspecification problem.
2. For this amount of estimated measurement error and proportion of censored observations the approximate proportional bias of the maximum likelihood estimator of α is 0.92 which equals the observed proportional bias, defined as the ratio of the maximum likelihood estimate to the GMM estimate of α.
3. In general the covariate coefficients corrections are, as expected, equal to correction in the shape parameter. The observed proportional bias for the statistically significant coefficients varies from 0.90 to 0.95.

[19]The second order properties of the test showed that this version of the test provides a reliable way of doing inference in the sense that the first order asymptotic distribution is a good approximation to the distribution of the test statistic. Also Monte Carlo experimentation showed that this version is more powerful than the Outer-Product-of-the-Grafient (OPG) version.

Figure 5: Residual analyses for Weibull and Log-logistic MLE estimates.



The influence of covariates is such that a positive coefficient accelerates the time to leaving unemployment, whereas a negative coefficient has the opposite effect. Apart from Age and Income in work, all coefficients are statistically significant at a 0.05 nominal level. The estimate results suggest the following comments:

1. The educational dummies have the expected positive effect.
2. Married (or leaving as a couple) individuals leave unemployment faster than single or divorced people.
3. Being the presence of dependent children in the household highly correlated with the level of benefits received from the government, it acts as a disincentive to return to the labour force.
4. Individuals in a high unemployment area exit unemployment slower as the rate of job offers is relatively lower and there is more competition for jobs.
5. The income variables have the expected sign. The higher the level of income in unemployment the lower the exit probability. Were the income in work variable statistically significant, the negative coefficient would indicate that those looking for jobs in higher wage jobs spend more time in unemployment[20].

An important feature of the GMM estimates is the little loss of efficiency comparatively to the maximum likelihood. Being a semiparametric estimator, there is always a trade-off between precision and flexibility. In this application the standard deviations of both estimates computed with the bootstrap have the same order of magnitude.

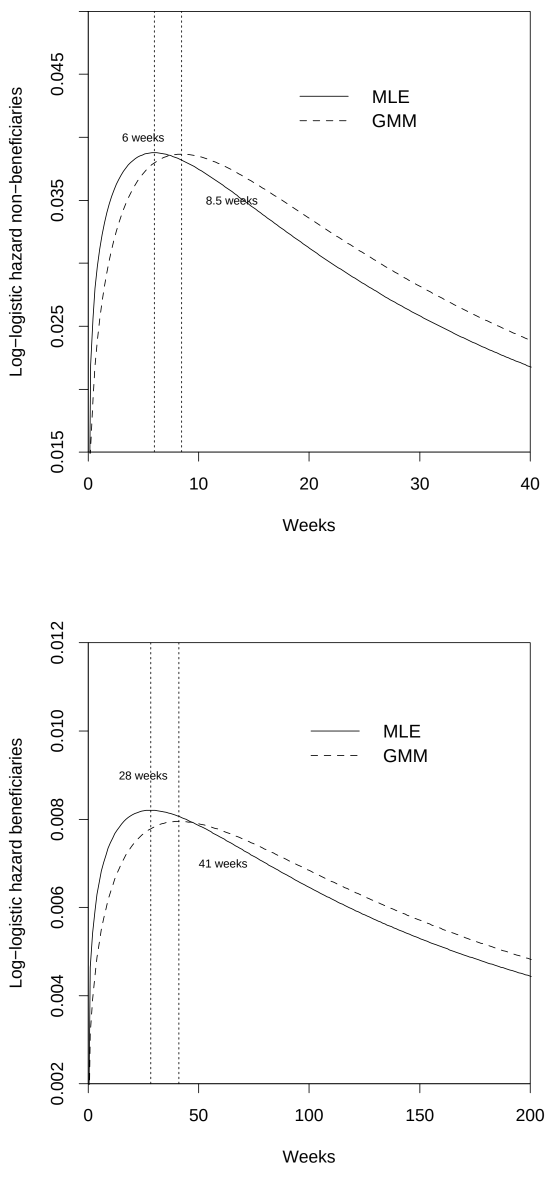
Figure 6 shows the conditional hazard functions for the two set of paramater estimates. Two cohorts are considered according to whether the individual receives any type of income support. The covariates are evaluated at the sample means of each cohort.

As predicted, the correction initially raises the hazard function above the MLE hazard. As an important consequence of this effect, the duration at which the Log-logistic hazard reaches its peak is smaller[21]. From the point of view of efficacy,

[20] In a previous version of the model, the expected wage at the job the individual is looking for was used, and for those who exit the state their current wages was assumed to be a realisation of that expectation. When this variable is used, the coefficient is significant and negative. Not only thit is a strong assumption, but also there is an endogeneity problem as this variable is clearly correlated with the reservation wage (see Nickell, 1979). On the other hand as noted in Lancaster and Chesher (1983) it is not straightforward to interpret thit variable as the mean wage or the conditional on being bigger than the reservation wage mean wage.

[21] The duration at which the Log-logistic hazard attains its maximum is given by $t_{(\max)} = [(\alpha - 1)\exp(-\bar{x}\beta)]^{1/\alpha}$, here evaluated at the GMM estimates and at the mean individual.

Figure 6: Estimated error contaminated (MLE) and bias corrected (GMM) hazard functions.



unemployment policies should target individuals on the increasing part of the hazard, as there it is easier to encourage them to return to work. As such, policies based on error contaminated durations may in this case loose its efficiency. Note that this issue is particularly relevant for beneficiaries, where the duration at which the MLE hazard reaches its maximum is nearly 50% larger.

These results assume that measurement error is homogenous. This is a strong assumption, as two spells with the same lenght that occurred in two different periods are likely to be contaminated with different amounts of measuremet error. The next section estimates the Log-logistic model assuming heteroskedastic measurement error of several forms.

6.4. Heteroskedastic measurement error. In most applications the distribution of duration response measurement error should be a function of the recall effort individuals have to make when reporting information on the spells. Given the nature of this data, it seems therefore natural to consider as a measure of recall effort (w) the sum of the time between the start of the spell and the date of interview with the time between the end of the spell and the date of interview. Note that this measure is independent of the spell lenght, as for example a short spell that happened a long time ago may have a larger recall effort than a large spell that just ended. Nevertheless large spells will always have a larger recall effort.

The variance of the measurement error will be a function of the logarithm of the recall effort. Four specifications for the skedastic function will be considered: the linear specification, which can be thought of as a first order local approximation and has the usual shortcome that may produce negative values for the variance; the exponential specification, which is always a natural candidate for skedastic functions; two picewise linear skedastic functions. In the first the threshold is the .75th quantile of w, and in the second the thresholds are the .25th and .75th quantiles of the distribution of w.

Table 8 shows the results for these specifications for the Log-logistic model[22]. The only specification that clearly rejects the skedastic function is the exponential. Its intercept estimates a measurement error variance of 0.303, which is very close to the homoskedastic model. The linear specification is not rejected at a 10% level, but when compared with the two picewise linear specifications is clearly rejected.

(Table 8 about here)

The preferred model is the picewise linear with two slopes. The correction on the shape parameter is now bigger, which shows that this model further identifies spurious variation on the log durations. On the other hand the estimates a less precise. The standard error of the shape parameter is twice as large as in the homoskedastic model. This may reflect the increased complexity associated with the introduction of

[22] The null hypothesis $H_0 : \pi_j = 0$ has a one sided alternative.

the skedastic function. The first slope is not significant, so up to the 75th quantile of the distribution of recall effort the measurement error variance is constant at 0.348, a value that is not to distant from the homoskedastic variance. It is after this quantile that a 1% variation on the recall effort induces an increase of the measurement error variance of 0,7. This is a very drastic increase in the variance and may not be solely related to measurement error. Remmember that extremely big values of the recall effort are likely to be associated with extremely long spells. In this case this specification could also be capturing the fact that those durations are not likely to be generated by a Log-logistic model.

7. Conclusion

This paper addressed the problem of GMM-Estimation under the presence of duration response measurement error. The impact on parameter estimates was characterized by deriving the approximate probability limit of estimators defined by a set of moment conditions. For single spell models, generally measurement error dampens the form of duration dependence in the hazard function. This effect differs from neglected uncontrolled heterogeneity, because the extent of the distortion is a function of the shape characteristics of the error-free distribution. In the cases here considered measurement error changes the way covariates affect the duration distribution, in the same fashion as it does for the shape parameters.

Allowing for right censoring and contaminated durations has different implications in different parametric specifications. In the Weibull model, right censoring offsets in an increasing way the negative impact of measurement error in MLE parameter estimates, while the approximate probability limit of the Log-logistic MLE estimates is a nonlinear function of the proportion of censored observations.

The seriousness of the implications of this misspecification problem are well illustrated in the two-spell-lagged-duration-dependence Exponential model. For this specification, estimated lagged duration dependence can be totally spurious. Depending on the sign of the correlation between the measurement errors, the magnitude and even the sign of this coefficient can be totally misperceived due to error-contamination.

This inconsistency measure leads to a GMM estimator that corrects (approximately) the bias in the moment conditions that define the error-free model. Its performance was investigated via Monte Carlo experimentation. The moment condition that defines the measurement error specification score test was shown to provide valuable information about the parameters of the distribution of the data under the presence of contamination, leading to the conclusion that under this estimating procedure score tests can be constructive. This estimator does not require any prior information on the measurement error distribution, whose parameters are estimated jointly with the parameters of the error free distribution.

The results were applied to a sample of unemployment durations retrospectively collected in the BHPS. The Weibull analysis suggests that parametric misspecification can give conflicting results between the maximum likelihood and the GMM estimates of the extended shape parameter vector. For the Log-logistic there was a very strong agreement between the MLE, and the GMM estimates. Introducing a measure of recall effort in the variance function raised the issue of independence between the measurement error and the spell length. The independence assumption may be too restrictive especially for large durations. On the other hand this parametrization of the variance function may just capture parametric misspecification.

If there is no measurement error this estimator can still be useful as it allows to answer the following question: "How much excessive variation in the log durations must one allow for when choosing for a given parametric specification". The estimate of the variance of measurement error can than be interpreted as the cost of choosing a parametric model. This estimator is particularly useful when the researcher has no information on the measurement error parameters or no access to external data to compute them.

A. The wage offer equation

In this appendix the mean of the wage distribution used as an explanatory variable in the specification of the unemployment duration model is estimated. The aim is to find a measure of the wage in the segment of the labour market in which the individual is searching for a job[23].

The wage offer equation is estimated using a standard Heckitt procedure like in Heckman (1979) which takes into account selection bias induced by observing wages only for employed people.

(Table 9 about here)

The data used was the sample of 3620 male individuals that were either employed or unemployed at time of interview of wave one.

In addition to the variables Children, Married and the educational dummies, the participation equation included a vector of explanatory variables measuring labour market experience (see Lambert, 1993 for a discussion on measures of labour market experience). The variable Experience is defined as the logarithim of the number of years since leaving full time education. The square of Experience was included to capture nonlinearities in the equation. The log wage equation included as explanatory variables, the educational dummies, the same experience measures and interactions with Age and the local unemployment rate.

[23]Other measures for this variable have been considered in this literature. Some authors use net earnings in previous job and others expected earnings at work. There are several reasons for not using those variables in the economic specification. The first is a practical one concerned with the size of the available sample: both the above variables are only available for a fraction of the sample considered in this study. The second reason is that, as noted in Nickell (1979), there is a potential endogeneity bias from using previous earnings, as those who are most likely to be selective about accepting jobs may have had higher than average earnings in their previous job. As for expected earnings, this variable is closely correlated with the reservation wage and with on job earnings.

Table 9 gives the sample descriptive statistics. Both employed and unemployed populations have very similar individual characteristics. However the latter seems to be younger, less experienced, less prone to being married but with more children.

Table 10 shows the results for both equations.(Table 10 about here)

B. THE APPROXIMATE STRUCTURAL BIAS FUNCTION FOR RIGHT CENSORED SINGLE SPELL MODELS

The approximate structural bias function for single spell models with right censored observations is now derived.

The aim is to find the functions $b_1^a(s,\theta)$ and $r_1(z,\theta)$, that solve the equation

$$\int_0^z g_1(s,\phi)m_T(s,\theta)ds = \int_0^z b_1^a(s,\theta)f_T(s,\phi)ds + r_1(z,\theta) \tag{32}$$

Using the definition of $m_T(s,\theta)$ the left hand side of (32) can be written as

$$\tfrac{\sigma^2}{2}\left(\int_0^z g_1(s,\phi)f_T\left(s,\phi\right)ds + 3\int_0^z g_1(s,\phi)sf_T'\left(s,\phi\right)ds + \int_0^z g_1(s,\phi)s^2 f_T''\left(s,\phi\right)ds\right) \tag{33}$$

Integrating the second term once by parts and the third term twice by parts, assuming the following tail conditions for the density and its partial derivatives, necessary to assure convergence of those integrals

$$\begin{array}{lll} A.B.1 & \lim_{s_k\to 0} g_1(s,\phi)sf_T(s,\phi) & = 0 \\ A.B.2 & \lim_{s_k\to 0} g_1(s,\phi)s^2 f_T'(s,\phi) & = 0 \\ A.B.3 & \lim_{s_k\to 0} g_1'(s,\phi)s^2 f_T(s,\phi) & = 0 \end{array} \tag{34}$$

leads to the desired result.

C. THE STRUCTURAL BIAS FUNCTION FOR MULTIPLE SPELLS SINGLE DESTINATION MODELS

In this appendix the structural bias function for MSSD models is derived which incorporates the case of SSSD if $R = 1$.

The approximation (5) to the multiple spell joint density of section 2 can be written as

$$\begin{aligned} f_{\mathbf{S}}(s) \simeq\ & f_{\mathbf{T}}(\mathbf{t}) + \Bigg\{\left(\tfrac{1}{2}\sum_{k=1}^{R}\sigma_k^2 + \sum_{k=1}^{R-1}\sum_{l=k+1}^{R}\sigma_{kl}\right)f_{\mathbf{T}}(\mathbf{s}) + \tfrac{3}{2}\sum_{k=1}^{R}\sigma_k^2 s_k f_{\mathbf{T}}^{(k)}(\mathbf{s}) + \\ & + \sum_{k=1}^{R}\sum_{l\neq k}^{R}\sigma_{kl}s_k f_{\mathbf{T}}^{(k)}(\mathbf{s}) + \tfrac{1}{2}\sum_{k=1}^{R}\sigma_k^2 s_k^2 f_{\mathbf{T}}^{(kk)}(\mathbf{s}) + \sum_{k=1}^{R-1}\sum_{l=k+1}^{R}\sigma_{kl}s_k s_l f_{\mathbf{T}}^{(kl)}(\mathbf{s})\Bigg\} \end{aligned} \tag{35}$$

where $f_{\mathbf{T}}^{(k)}\left(\mathbf{s}\right) = \partial f_{\mathbf{T}}\left(\mathbf{s}\right)/\partial t_k$ and $f_{\mathbf{T}}^{(kl)}\left(\mathbf{s}\right) = \partial^2 f_{\mathbf{T}}\left(\mathbf{s}\right)/\partial t_k \partial t_l$.

Computation of $E_{\mathbf{S}}[\mathbf{g}(\mathbf{S},\boldsymbol{\phi})|\boldsymbol{\theta} = \boldsymbol{\theta}]$, up to $o(||\boldsymbol{\sigma}||)$ requires calculation of three integrals,

$$
\begin{array}{lll}
1. & \int_0^\infty \ldots \int_0^\infty \mathbf{g}(\mathbf{s},\boldsymbol{\phi})\, s_k f_{\mathbf{T}}^{(k)}(\mathbf{s})\, ds_R \ldots ds_1 & = -(E_{\boldsymbol{\phi}}[\mathbf{g}(\mathbf{S},\boldsymbol{\phi})] + E_{\boldsymbol{\phi}}[S_k \mathbf{g}^{(k)}(\mathbf{S},\boldsymbol{\phi})]) \\
2. & \int_0^\infty \ldots \int_0^\infty \mathbf{g}(\mathbf{s},\boldsymbol{\phi})\, s_k^2 f_{\mathbf{T}}^{(kk)}(\mathbf{s})\, ds_R \ldots ds_1 & = 2E_{\boldsymbol{\phi}}[\mathbf{g}(\mathbf{S},\boldsymbol{\phi})] + 4E_{\boldsymbol{\phi}}[S_k \mathbf{g}^{(k)}(\mathbf{S},\boldsymbol{\phi})] + \\
 & & + E_{\boldsymbol{\phi}}[S_k^2 \mathbf{g}^{(kk)}(\mathbf{S},\boldsymbol{\phi})] \\
3. & \int_0^\infty \ldots \int_0^\infty \mathbf{g}(\mathbf{s},\boldsymbol{\phi})\, s_k s_l f_{\mathbf{T}}^{(kl)}(\mathbf{s})\, ds_R \ldots ds_1 & = -(E_{\boldsymbol{\phi}}[\mathbf{g}(\mathbf{S},\boldsymbol{\phi})] + E_{\boldsymbol{\phi}}[S_k \mathbf{g}^{(k)}(\mathbf{S},\boldsymbol{\phi})] + \\
 & & + E_{\boldsymbol{\phi}}[S_l \mathbf{g}^{(l)}(\mathbf{S},\boldsymbol{\phi})] + E_{\boldsymbol{\phi}}[S_k S_l \mathbf{g}^{(kl)}(\mathbf{S},\boldsymbol{\phi})])
\end{array}
\tag{36}
$$

Computation of those integrals required multiple integration by parts, and assumption of the following conditions related to the tail behaviour of the density and its partial derivatives,

$$
\begin{array}{llllll}
A.C.1 & \lim_{s_k \to 0} \mathbf{g}(\mathbf{s},\boldsymbol{\phi}) s_k f_{\mathbf{T}}(\mathbf{s}) & = & \lim_{s_k \to \infty} \mathbf{g}(\mathbf{s},\boldsymbol{\phi}) s_k f_{\mathbf{T}}(\mathbf{s}) & = & 0 \\
A.C.2 & \lim_{s_k \to 0} \mathbf{g}(\mathbf{s},\boldsymbol{\phi}) s_k^2 f_{\mathbf{T}}^{(k)}(\mathbf{s}) & = & \lim_{s_k \to \infty} \mathbf{g}(\mathbf{s},\boldsymbol{\phi}) s_k^2 f_{\mathbf{T}}^{(k)}(\mathbf{s}) & = & 0 \\
A.C.3 & \lim_{s_k \to 0} \mathbf{g}(\mathbf{s},\boldsymbol{\phi}) s_k s_l f_{\mathbf{T}}^{(l)}(\mathbf{s}) & = & \lim_{s_k \to \infty} \mathbf{g}(\mathbf{s},\boldsymbol{\phi}) s_k s_l f_{\mathbf{T}}^{(l)}(\mathbf{s}) & = & 0 \\
A.C.4 & \lim_{s_l \to 0} \mathbf{g}^{(k)}(\mathbf{s},\boldsymbol{\phi}) s_k s_l f_{\mathbf{T}}(\mathbf{s}) & = & \lim_{s_l \to \infty} \mathbf{g}^{(k)}(\mathbf{s},\boldsymbol{\phi}) s_k s_l f_{\mathbf{T}}(\mathbf{s}) & = & 0
\end{array}
\tag{37}
$$

The approximate required expectation can now be written as

$$
\begin{aligned}
E_{\mathbf{S}}[\mathbf{g}(\mathbf{S},\boldsymbol{\phi})|\boldsymbol{\theta}=\boldsymbol{\theta}] &\simeq a_1(\boldsymbol{\sigma})E_{\mathbf{T}}[\mathbf{g}(\mathbf{S},\boldsymbol{\phi})|\boldsymbol{\phi}=\boldsymbol{\phi}] + a_2(\boldsymbol{\sigma})E_{\mathbf{T}}[S_k \mathbf{g}^{(k)}(\mathbf{S},\boldsymbol{\phi})|\boldsymbol{\phi}=\boldsymbol{\phi}] \\
&+ a_3(\boldsymbol{\sigma})E_{\mathbf{T}}[S_k^2 \mathbf{g}^{(kk)}(\mathbf{S},\boldsymbol{\phi})|\boldsymbol{\phi}=\boldsymbol{\phi}] + a_4(\boldsymbol{\sigma})E_{\mathbf{T}}[S_k S_l \mathbf{g}^{(kl)}(\mathbf{S},\boldsymbol{\phi})|\boldsymbol{\phi}=\boldsymbol{\phi}]
\end{aligned}
\tag{38}
$$

where $a_j(\boldsymbol{\sigma})$, $j = 1, ..., 4$ are polynomial functions of the vector $\boldsymbol{\sigma}$.

1. The coefficient of $E_{\mathbf{T}}[\mathbf{g}(\mathbf{S},\boldsymbol{\phi})|\boldsymbol{\phi}=\boldsymbol{\phi}]$ is

$$
\begin{aligned}
a_1(\boldsymbol{\sigma}) &= -\frac{3}{2}\sum_{k=1}^{R}\sigma_k^2 - \sum_{k=1}^{R}\sum_{l\neq k}^{R}\sigma_{kl} + \sum_{k=1}^{R}\sigma_k^2 + \frac{1}{2}\sum_{k=1}^{R}\sigma_k^2 + 2\sum_{k=1}^{R-1}\sum_{l=k+1}^{R}\sigma_{kl} \\
&= -\sum_{k=1}^{R}\sum_{l\neq k}^{R}\sigma_{kl} + 2\sum_{k=1}^{R-1}\sum_{l=k+1}^{R}\sigma_{kl} = 0
\end{aligned}
\tag{39}
$$

2. The coefficient of $E_{\mathbf{T}}[s_k \mathbf{g}^{(k)}(\mathbf{s},\boldsymbol{\phi})|\boldsymbol{\phi}=\boldsymbol{\phi}] \equiv \bar{w}_k(\boldsymbol{\phi})$ is

$$
\begin{aligned}
a_2(\boldsymbol{\sigma}) &= -\frac{3}{2}\sum_{k=1}^{R}\sigma_k^2\bar{w}_k(\boldsymbol{\phi}) - \sum_{k=1}^{R}\sum_{l\neq k}^{R}\sigma_{kl}\bar{w}_k(\boldsymbol{\phi}) + 2\sum_{k=1}^{R}\sigma_k^2\bar{w}_k(\boldsymbol{\phi}) + \\
&+ \sum_{k=1}^{R-1}\sum_{l=k+1}^{R}\sigma_{kl}\bar{w}_k(\boldsymbol{\phi}) + \sum_{k=1}^{R-1}\sum_{l=k+1}^{R}\sigma_{kl}\bar{w}_l(\boldsymbol{\phi})
\end{aligned}
\tag{40}
$$

Noting that

$$\sum_{k=1}^{R}\sum_{l\neq k}^{R}\sigma_{kl}\bar{w}_k(\boldsymbol{\phi}) = \sum_{k=1}^{R-1}\sum_{l=k+1}^{R}\sigma_{kl}\bar{w}_k(\boldsymbol{\phi}) + \sum_{k=1}^{R-1}\sum_{l=k+1}^{R}\sigma_{kl}\bar{w}_l(\boldsymbol{\phi}) \tag{41}$$

gives

$$a_2(\boldsymbol{\sigma}) = \frac{1}{2}\sum_{k=1}^{R}\sigma_k^2\bar{w}_k(\boldsymbol{\phi}) \tag{42}$$

3. Finally the terms associated with

$$E_{\mathbf{T}}[S_k^2\mathbf{g}^{(kk)}(\mathbf{S},\boldsymbol{\phi})|\boldsymbol{\phi}=\boldsymbol{\phi}] \equiv \bar{w}_{kk}(\boldsymbol{\phi}); \qquad E_{\mathbf{T}}[S_kS_l\mathbf{g}^{(kl)}(\mathbf{S},\boldsymbol{\phi})|\boldsymbol{\phi}=\boldsymbol{\phi}] \equiv \bar{w}_{kl}(\boldsymbol{\phi}), \tag{43}$$

are respectively

$$a_3(\boldsymbol{\sigma}) = \frac{1}{2}\sum_{k=1}^{R}\sigma_k^2\bar{w}_{kk}(\boldsymbol{\phi}); \qquad a_4(\boldsymbol{\sigma}) = \sum_{k=1}^{R-1}\sum_{l=k+1}^{R}\sigma_{kl}\bar{w}_{kl}(\boldsymbol{\phi}) \tag{44}$$

It follows that the approximate structural bias function is given by

$$\mathbf{b}^a(\mathbf{s},\boldsymbol{\theta}) = \frac{1}{2}\sum_{k=1}^{R}\sigma_k^2\left(s_k\mathbf{g}^{(k)}(\mathbf{s},\boldsymbol{\phi}) + s_k^2\mathbf{g}^{(kk)}(\mathbf{s},\boldsymbol{\phi})\right) + \sum_{k=1}^{R-1}\sum_{l=k+1}^{R}\sigma_{kl}s_ks_l\mathbf{g}^{(kl)}(\mathbf{s},\boldsymbol{\phi}). \tag{45}$$

D. SIGMA-ORDER CONSISTENCY OF THE APPROXIMATE GMM ESTIMATOR

In this appendix the σ order of the bias corrected GMM estimator is derived.

Let Y be a random variable whose distribution depends on $\theta = \{\phi, \sigma\}$. Let θ_0 be the true value, and consider the class of extremum estimators in which under the condition that $E[g(Y,\theta_0)|\theta = \theta_0] = 0$, estimators are obtained by maximizing an approximation to a true objective function, $\widehat{Q}_n(\theta) = -\hat{g}_n(\theta)'\hat{W}\hat{g}_n(\theta)$, where $\hat{g}_n(\theta) = n^{-1/2}\sum_{i=1}^{n} g(y_i,\theta)$. By the law of large numbers $\hat{g}_n(\theta) \xrightarrow{p} g_0(\theta) = E[g(Y,\theta)|\theta = \theta_0]$. Then by a continuity argument $\widehat{Q}_n(\theta) \xrightarrow{p} Q_0(\theta) = -g_0(\theta)'Wg_0(\theta)$, is the probability limit of the true objective function, and convergence in probability is uniform.

The objective function $\widehat{Q}^a_n(\theta) = -\hat{g}^a_n(\theta)'\hat{W}\hat{g}^a_n(\theta)$, maximized at $\hat{\theta}^a_n$, is obtained by approximating the influence of a subset of parameters σ, on the moment conditions in a way that

$$E[g^a(Y,\theta_0)|\theta = \theta_0] = O\left(\sigma_0^3\right) \tag{46}$$

Assuming that $\hat{g}^a_n(\theta) \xrightarrow{p} g^a_0(\theta) = E[g^a(Y,\theta)|\theta = \theta_0]$, the probability limit of the approximate objective function is $\widehat{Q}^a_n(\theta) \xrightarrow{p} Q^a_0(\theta) = g^a_0(\theta)'Wg^a_0(\theta)$.

Theorem 1. *Let $Q_0(\theta)$ be the probability limit of the true objective function, and let θ_0 be the true value of θ assumed identifiable in the sense that*

$$\theta_0 = \arg\max_{\theta} Q_0(\theta) = -g_0(\theta)' W g_0(\theta) \tag{47}$$

defines an unique value of θ_0. Let θ^a the probability limit of the approximate estimator be uniquely defined by

$$\theta^a = \arg\max_{\theta} Q_0^a(\theta) = -g_0^a(\theta)' W g_0^a(\theta) \tag{48}$$

Then $\theta^a - \theta_0 = O\left(\sigma_0^3\right)$.

The proof exploits the fact that $\hat{\theta}_n^a$ has an influence function representation (see Newey and McFadden, 1994) and that its distribution is degenerate, therefore convergence in distribution implies convergence in probability.

Assume θ_0 is in the interior of its parameter space Θ.The first order condition for $\hat{\theta}_n^a$ has the form,

$$G_n^a(\hat{\theta}_n^a) W g_n^a(\hat{\theta}_n^a) = 0 \tag{49}$$

where $G_n^a(\theta) = \nabla_\theta g_n^a(\theta)$. Assume that $g^a(y,\theta)$ is continuously differentiable on int(Θ). A mean value expansion of $g_n^a(\hat{\theta}_n^a)$ about θ_0 gives

$$G_n^a(\hat{\theta}_n^a)' W [g_n^a(\theta_0) + G_n^a(\ddot{\theta}_n)(\hat{\theta}_n^a - \theta_0)] = 0 \tag{50}$$

where $\ddot{\theta}_n$ is between $\hat{\theta}_n^a$ and θ_0. Therefore,

$$n^{1/2}(\hat{\theta}_n^a - \theta_0) = -[G_n^a(\hat{\theta}_n^a)' W G_n^a(\ddot{\theta}_n)]^{-1} G_n^a(\hat{\theta}_n^a) W n^{1/2}\ g_n^a(\theta_0) \tag{51}$$

Because $\hat{\theta}_n^a \xrightarrow{p} \theta^a$ and θ_0 is the true parameter vector, under standard regularity conditions $G_n^a(\hat{\theta}_n^a) \xrightarrow{p} G_\theta^a$ and $G_n^a(\ddot{\theta}_n) \xrightarrow{p} G_\theta^a$ where $G_\theta^a = E[\nabla_\theta g^a(Y,\theta^a)|\theta = \theta_0]$. Let $A = -(G_\theta^{a\prime} W G_\theta^a)^{-1} G_\theta^{a\prime} W$ and write

$$n^{1/2}(\hat{\theta}_n^a - \theta_0) = A\ n^{1/2}(g_n^a(\theta_0) - O(\sigma_0^3)) + n^{1/2} O(\sigma_0^3) + o_p(1) \tag{52}$$

or equivalently

$$n^{1/2}(\hat{\theta}_n^a - \theta_0 - O(\sigma_0^3)) = A n^{1/2}(g_n^a(\theta_0) - O(\sigma_0^3)) + o_p(1) \tag{53}$$

It follows directly from (46) that $n^{1/2}(g_n^a(\theta_0) - O(\sigma_0^3)) \xrightarrow{d} N(0,V)$ where

$$V = E[(g^a(Y,\theta_0) - O(\sigma_0^3))(g^a(Y,\theta_0) - O(\sigma_0^3))'|\theta = \theta_0] \tag{54}$$

and that implies $g_n^a(\theta_0) - O(\sigma_0^3) \xrightarrow{p} 0$. As a direct consequence,

$$\hat{\theta}_n^a - \theta_0 \xrightarrow{p} O(\sigma_0^3). \tag{55}$$

from which it follows that $\theta^a - \theta_0 = O\left(\sigma_0^3\right)$.

Acknowledgements

I am grateful to Andrew Chesher, João Santos Silva, João Manuel Andrade e Silva, Arnstein Aassve and Nicoletta Rosati for helpful suggestions. I am in debt to Paris Nearchou for providing the data for the application. Symbolic computation was conducted using Mathematica (Wolfram, 1992); Monte Carlo experiments were conducted using Splus (Statistical Sciences, 1993.). This study was supported by Fundação para a Ciência e Tecnologia, program POCTI, partially funded by FEDER..

REFERENCES

[1] ABREVAYA, J., AND J.A. HAUSMAN (1999): "Semiparametric estimation with mismeasured dependent variables: An application to duration models for unemployment spells," *Annales d'économie et de statistique*, No.55-56, 243-276.

[2] ATKINSON, A.B., AND J. FLEMMING (1978): "Unemployment, social security and Incentives," *Midland Bank Review.*

[3] ATKINSON, A.B., GOMULKA, J., MICKELWRIGHT, J., AND J. FLEMMING (1982): "Unemployment benefits, duration, and incentives in Britain: How robust is the evidence?," International Centre for Economic and Related Disciplines, Discussion paper No. 43.

[4] BUONACCORSI, J.P. (1996): "A modified estimating equation approach to correcting for measurement error in regression," *Biometrika*, 83, 433-440.

[5] BUZAS, J.S., AND L.A. STEFANSKI (1995): "A note on corrected score estimation," *Statistics & Probability Letters*, 28, 1-18.

[6] BRENDAN, H. (1997): "Unified BHPS work-life histories: Combining multiple sources into a user-friendly format," *Technical Papers of the ESRC Research Centre on Micro-social Change.* Technical Paper 13. Colchester: University of Essex.

[7] CARROLL, R.J. AND L.A. STEFANSKI (1990): "Approximate Quasilikelihood estimation in models with surrogate predictors," *Journal of the American Statistical Association*, 85, 652-663.

[8] CAVANAGH, C., AND R.P. SHERMAN (1998): "Rank estimators for Monotonic Index models," *Journal of Econometrics*, 84, 351-381.

[9] CHESHER, A.D. (1991) "The effect of measurement error," *Biometrika*, 78, 451-462.

[10] CHESHER, A.D. (1998) "Measurement error bias reductions," Discussion paper No. 98/449, Department of Economics, University of Bristol.

[11] CHESHER, A.D. (1999) "Structural estimation in the presence of covariate measurement error," revision of Chesher, A.D. (1998), unpublished.

[12] CHESHER, A.D. (2000) "Improved GMM estimation under covariate measurement error," presented at the World Meeting of the Econometric Society, Seattle.

[13] CHESHER, A.D. (2001) "Parameter approximations for quantile regressions with measurement error," Cemmap working paper CWP02/01.

[14] CHESHER, A.D., DUMANGANE, M., AND R.J. SMITH (2002): "Duration response measurement error," *Journal of Econometrics*, 111, 169-194.

[15] CHESHER, A.D., AND J.M.S. SILVA (2002): "Taste variation in Discrete Choice models," *Review of Economics Studies*, 69, 147-168.

[16] COX, D.R. (1972): *Renewal Theory.* London: Metheun.

[17] DEX, S., AND A. MCCULLOCH (1997): "The reliability of retrospective unemployment history data," Unpublished paper, Judge Institute for Management Studies, Cambridge.

[18] DUMANGANE, M. (2000): *"Essays on Duration Response Measurement Error"*. PhD Thesis. University of Bristol.

[19] ELBERS, C., AND G. RIDDER (1982): "True and spurious duration dependence: The identifiability of the Proportional Hazard model," *Review of Economic Studies*, 49, 403-409.

[20] ELIAS, P. (1996): "Who forgot they were unemployed?," Unpublished paper, Institute for Employment Research, University of Warwick.

[21] HECKMAN, J.J. (1979): "Sample selection bias as a specification error," *Econometrica*, 47, 153-161.

[22] HECKMAN, J.J. (1991): "Identifying the hand of past: distinguishing state dependence from heterogeneity," *American Economic Review*, 81, 75-79.

[23] 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.

[24] HECKMAN, J.J., AND C.R. TABER (1994): "Econometric mixture models and more general models for unobservables in duration analysis," *NBER*, technical working paper No.157.

[25] HOROWITZ, J.L. (1999): "Semiparametric estimation of a Proportional Hazard model with unobserved heterogeneity," *Econometrica*, 67, 1001-1028.

[26] Kalbfleisch, J.D. and R.L. Prentice. (1980): *The Statistical Analysis of Failure Time Data*. New York: Wiley.

[27] Kiefer, N.M. and G.R. Skoog (1988): "Local asymptotic specification error analysis," *Econometrica,* 52, 873-885.

[28] Lambert, S.F. (1993): "Labour market experience in female wage equations: Does the experience measure matter?," *Applied Economics*, 25, 1439-1449.

[29] Lancaster, T. (1979): "Econometric methods for the duration of unemployment," *Econometrica*, 47, 939-956.

[30] Lancaster, T. (1990): *The Econometrics of Transition Data*. Cambridge: Cambridge University Press.

[31] Lancaster, T., and A.D. Chesher (1983): "An econometric analysis of reservation wages," *Econometrica*, 51, 1661-1676.

[32] Lancaster, T., and A.D. Chesher (1985): "Residual analysis for censored duration data," *Economics Letters*, 12, 723-725.

[33] Lancaster, T., and S. Nickell (1980): "The analysis of reemployment probabilities for the unemployed," *Journal of the Royal Statistical Society* A, 143, 141-165.

[34] Lawless, J.F. (1982): *Statistical Methods and Methods for Lifetime Data*. New York: Wiley.

[35] McCullagh, P.N., and J.A. Nelder (1989): *Generalized Linear Models, second edition*. London: Chapman & Hall.

[36] Nakamura, T. (1990) "Corrected score functions for errors-in-variables models: Methodology and application to Generalized Linear models," *Biometrika*, 77, 127-137.

[37] Narendranathan, W., Nickell, S., and J. Stern (1985): "Unemployment benefits revisited," *Economic Journal*, 95, 307-329.

[38] Newey, W.K., and D.L. McFadden (1994): "Large sample estimation and hypothesis testing," in *Hanbook of Econometrics, Volume 4,* edited by Engle, R.F., and D.L. McFadden. Amsterdam: Elsevier Science B.V.

[39] Nickell, S. (1979a): "The effect of unemployment and related benefits on the duration of unemployment," *Economic Journal*, 89, 34-49.

[40] Nickell, S. (1979b): "Estimating the probability of leaving unemployment," *Econometrica*, 47, 1249-1266.

[41] PACE, L., AND A. SALVAN (1997): "*Principles of Statistical Inference From a Neo-Fisherian Perspective*" . London: World Scientific.

[42] PAULL, G., (1997): "Dynamic labour market behaviour in the British Household Panel survey: The effects of recall bias and panel attrition," Centre for Economic Research & Institute of Economics 10.

[43] PAULL, G., (2000): "Biases in the reporting of labour market dynamics," Working Paper W02/10, Institute for Fiscal Studies.

[44] ROMEO, C.J. (1997): "Measuring information loss due to inconsistencies in duration data from longitudinal surveys," *Journal of Econometrics*, 78, 159-177.

[45] SKINNER, C.J., AND K. HUMPHREYS (1999) "Weibull regression for lifetimes measured with error," *Lifetime Data Analysis*, 5, 23-37.

[46] STATISTICAL SCIENCES (1993): *S-PLUS® User's Manual, Version 3.2.* Seattle: StatSci, a division of MathSoft Inc.

[47] STEFANSKI, L.A. (1989) "Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models," *Communications in Statistics, Series A*, 18, 4335-4358.

[48] STEFANSKI, L.A., AND R.J. CARROLL (1987) "Conditional and optimal scores in Generalized Linear measurement error models," *Biometrika*, 74, 703-716.

[49] WOLFRAM, S. (1992): Mathematica*, a system for doing mathematics by computer, 2nd edition,* Reading, Mass: Addison-Wesley.

[50] WOLTER, K.M., AND W.A. FULLER (1982): "Estimation of nonlinear errors-in-variables models," *Annals of Statistics*, 10, 539-548.

Table 1: Weibull and Log-logistic GMM bias corrected estimates under no measurement error.

		Weibull											
		β_0			β_1			α			σ^2		
δ	N	mean	med	sd	mean	med	sd	mean	med	sd	mean	med	sd
0.20	200	1.014	1.005	.124	1.008	1.000	.164	1.010	1.005	.083	-.022	-.017	.080
	500	1.008	1.003	.085	1.004	1.002	.105	1.006	1.002	.057	-.010	-.006	.060
0.50	200	1.029	1.006	.206	1.014	.996	.219	1.015	1.003	.113	-.043	-.026	.144
	500	1.026	1.010	.140	1.013	1.006	.140	1.014	1.005	.080	-.019	-.011	.114
		Log-logistic											
0.20	200	1.022	1.004	.180	1.020	1.001	.265	1.024	1.004	.136	.013	.000	.426
	500	1.016	1.009	.119	1.012	1.004	.167	1.021	1.007	.095	.045	.000	.345
0.50	200	1.067	1.018	.307	1.051	1.012	.335	1.054	1.008	.200	.052	.004	.521
	500	1.049	1.018	.220	1.041	1.016	.229	1.041	1.007	.154	.072	-.004	.424

Table 2: Weibull maximum likelihood estimates ignoring measurement error.

				β_0			β_1			α		
δ	ρ	N	V	mean	med	sd	mean	med	sd	mean	med	sd
0.20	0.90	200	LN	.889	.886	.099	.926	.923	.153	.928	.925	.058
			G	.896	.895	.099	.934	.931	.151	.934	.932	.058
		500	LN	.879	.877	.062	.922	.920	.094	.921	.919	.037
			G	.887	.886	.061	.929	.928	.095	.926	.925	.037
	0.80	200	LN	.747	.743	.101	.836	.838	.155	.842	.839	.053
			G	.772	.770	.097	.856	.850	.148	.859	.857	.051
		500	LN	.738	.738	.063	.837	.837	.101	.835	.835	.034
			G	.762	.761	.060	.854	.853	.098	.853	.851	.033
0.50	0.90	200	LN	.903	.904	.142	.944	.936	.197	.947	.946	.073
			G	.913	.912	.144	.951	.944	.194	.953	.949	.073
		500	LN	.889	.885	.087	.939	.939	.117	.937	.935	.045
			G	.901	.899	.084	.946	.944	.119	.943	.941	.046
	0.80	200	LN	.712	.711	.138	.849	.847	.189	.856	.853	.064
			G	.748	.748	.134	.871	.871	.186	.875	.873	.064
		500	LN	.700	.700	.086	.849	.848	.123	.848	.848	.041
			G	.734	.735	.082	.866	.868	.120	.867	.866	.040

Table 3: Log-logistic maximum likelihood estimates ignoring measurement error.

				β_0			β_0			α		
δ	ρ	N	V	mean	med	sd	mean	med	sd	mean	med	sd
0.20	0.90	200	LN	.955	.953	.146	.950	.949	.228	.954	.951	.063
			G	.950	.947	.145	.952	.950	.228	.955	.951	.062
		500	LN	.942	.939	.092	.941	.942	.144	.943	.942	.038
			G	.940	.940	.091	.942	.942	.143	.945	.945	.039
	0.80	200	LN	.860	.861	.142	.884	.885	.227	.881	.879	.056
			G	.849	.845	.141	.886	.882	.226	.884	.881	.056
		500	LN	.853	.855	.092	.875	.875	.140	.876	.875	.036
			G	.843	.839	.088	.878	.878	.145	.879	.878	.036
0.50	0.90	200	LN	.925	.927	.187	.946	.942	.259	.951	.946	.080
			G	.925	.923	.186	.951	.949	.257	.954	.951	.079
		500	LN	.911	.911	.117	.938	.937	.165	.939	.939	.046
			G	.912	.911	.115	.940	.936	.164	.942	.942	.047
	0.80	200	LN	.761	.753	.173	.868	.869	.260	.865	.864	.067
			G	.763	.757	.172	.874	.864	.257	.874	.871	.069
		500	LN	.753	.754	.114	.860	.855	.160	.860	.857	.043
			G	.757	.757	.110	.869	.865	.165	.868	.866	.044

Table 4: Weibull GMM bias corrected estimates under measurement error.

				β_0			β_0			α			σ^2		
δ	ρ	N	V	mean	med	sd	mean	med	sd	mean	med	sd	mean	med	sd
0.20	0.90	200	LN	.987	.988	.137	.994	.986	.197	1.006	.996	.088	.106	.085	.153
			G	.989	.981	.131	1.006	1.002	.183	1.010	1.000	.087	.098	.078	.140
		500	LN	.968	.961	.102	.984	.974	.126	.993	.983	.068	.102	.088	.107
			G	.968	.962	.104	.993	.981	.133	.996	.984	.070	.089	.080	.095
	0.80	200	LN	.832	.840	.219	.921	.924	.224	.963	.939	.142	.330	.140	.411
			G	.837	.838	.212	.947	.947	.222	.978	.947	.141	.329	.131	.432
		500	LN	.772	.797	.198	.939	.914	.180	.939	.914	.118	.320	.128	.370
			G	.774	.802	.184	.942	.931	.165	.953	.922	.114	.312	.120	.356
0.50	0.90	200	LN	1.024	1.014	.165	1.020	1.009	.220	1.018	1.011	.090	.101	.087	.163
			G	1.031	1.024	.164	1.025	1.015	.217	1.023	1.014	.087	.095	.088	.123
		500	LN	.997	.993	.119	1.002	1.003	.136	.1001	.999	.064	.104	.010	.113
			G	1.000	.993	.112	1.008	1.006	.137	1.002	.997	.062	.087	.086	.092
	0.80	200	LN	.840	.847	.228	.928	.922	.242	.946	.938	.121	.337	.162	.466
			G	.861	.857	.228	.950	.938	.241	.963	.951	.126	.345	.164	.488
		500	LN	.790	.804	.180	.910	.907	.171	.917	.916	.093	.318	.148	.415
			G	.813	.819	.192	.928	.923	.174	.938	.930	.107	.346	.161	.442

Table 5: Log-logistic GMM bias corrected estimates under measurement error

				β_0			β_1			α			σ^2		
δ	ρ	N	V	mean	med	sd	mean	med	sd	mean	med	sd	mean	med	sd
0.20	0.90	200	LN	1.001	.987	.180	1.007	.990	.266	1.016	.995	.126	.263	.196	.330
			G	.994	.984	.172	1.008	.996	.260	1.015	.995	.120	.252	.164	.326
		500	LN	.986	.981	.113	.996	.994	.164	1.003	.995	.078	.285	.333	.251
			G	.980	.976	.107	.994	.990	.161	1.001	.991	.077	.266	.321	.245
	0.80	200	LN	.915	.906	.172	.956	.943	.268	.961	.940	.120	.415	.316	.419
			G	.907	.894	.179	.964	.946	.266	.968	.942	.129	.424	.307	.444
		500	LN	.917	.915	.113	.960	.955	.169	.969	.966	.082	.528	.630	.340
			G	.902	.897	.109	.962	.956	.173	.968	.965	.084	.496	.575	.354
0.50	0.90	200	LN	1.010	.977	.282	1.019	.998	.316	1.028	.996	.164	.297	.236	.376
			G	1.002	.969	.263	1.023	1.001	.309	1.027	.998	.159	.283	.169	.376
		500	LN	.995	.962	.216	1.014	.993	.224	1.016	.986	.140	.315	.327	.339
			G	.985	.958	.194	1.009	.994	.213	1.010	.984	.128	.282	.291	.345
	0.80	200	LN	.845	.823	.235	.947	.932	.308	.952	.928	.138	.471	.370	.477
			G	.828	.811	.209	.946	.927	.289	.945	.929	.116	.399	.202	.452
		500	LN	.840	.829	.155	.949	.942	.197	.955	.947	.095	.577	.783	.398
			G	.823	.815	.143	.943	.932	.194	.942	.928	.092	.436	.343	.404

Table 6: Summary statistics for unemployment duration model.

	All		Benef.		Non Benef.	
Number of spells	510	—	403	—	107	—
Censored spells	0.60	—	0.65	—	0.41	—
Uncensored spell length (in weeks)	27.4	(47.2)	32.6	(55.3)	15.8	(13.7)
Censored spell length (in weeks)	64.4	(99.0)	69.9	(100.1)	31.8	(86.6)
Age	33.6	(14.0)	34.7	(14.0)	29.5	(13.2)
Higher education	0.23	—	0.22	—	0.30	—
Lower education	0.42	—	0.41	—	0.47	—
Married	0.56	—	0.59	—	0.44	—
Children	0.56	(0.7)	0.56	(0.8)	0.59	(0.7)
Local unemployment rate (in %)	7.8	(1.6)	7.9	(1.6)	7.5	(1.5)
Income in unemployment (£ per week)	53.5	(55.5)	67.6	(54.2)	—	—
Income in work (£ per week)	160.1	(52.7)	161.0	(48.3)	156.9	(66.8)

* Standard errors in parentheses for continuous variables

Table 7: Weibull and Log-logistic MLE and GMM corrected estimates

	Weibull				Log-logistic			
Variable	MLE		GMM		MLE		GMM	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Alpha	0.880*	.057	0.946	.056	1.232*	.069	1.333*	.073
Constant	-3.326*	.195	-3.337*	.201	-3.840*	.247	-4.132*	.231
Age	-0.892*	.394	-0.370	.280	-0.451	.398	-0.483	.401
Higher education	1.322*	.282	1.538*	.270	1.713*	.342	1.798*	.328
Low education	0.702*	.284	1.014*	.244	1.119*	.339	1.215*	.334
Married	0.636*	.199	0.729*	.196	0.862*	.256	0.938*	.264
Children	-0.372*	.135	-0.476*	.137	-0.506*	.158	-0.564*	.166
Unemployment rate	-0.101**	.055	-0.086	.058	-0.148*	.070	-0.164*	.075
Beneficiary	-0.784*	.258	-1.116*	.200	-1.203*	.235	-1.346*	.256
Income in unemployment	-0.423*	.059	-0.431*	.063	-0.612*	.099	-0.653*	.112
Income in work (B)	-0.269	.354	-0.591**	.305	-0.488	.438	-0.482	.446
Income in work (N. B.)	-0.237	.391	-0.608*	.288	-0.369	.345	-0.343	.322
ME variance	—	—	0.003	.031	—	—	0.306*	.173
	ME Test=83.75				ME Test=1.26			

* Rejected at 5%

**Rejected at 10%

Table 8: Log-logistic GMM corrected estimates with heteroskedastic measurement error

Variable	Linear		Exponential		Picewise1		Picewise2	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Alpha	1.348*	.086	1.396	.176	1.435*	.151	1.422*	.155
Constant	-4.133*	.258	-4.274	.459	-4.390*	.382	-4.345*	.414
Age	-0.494	.420	-0.485	.496	-0.386	.486	-0.345	.450
Higher education	1.828*	.350	1.880	.415	-1.937*	.392	1.939*	.378
Low education	1.265*	.361	1.335	.436	1.392*	.397	1.389*	.380
Married	0.979*	.277	1.015	.327	1.039*	.316	1.031*	.301
Children	-0.600*	.175	-0.632	.211	-0.660*	.208	-0.654*	.203
Local unemployment rate	-0.177*	.077	-0.186	.088	-0.198*	.085	-0.195*	.082
Beneficiary	-1.399*	.261	-1.453	.315	-1.481	.317	-1.469*	.297
Income in unemployment	-0.652*	.114	-0.670	.133	-0.685*	.132	-0.681*	.125
Income in work (Benef)	-0.488	.455	-0.488	.522	-0.560	.503	-0.580	.484
Income in work (Non benef)	-0.331	.332	-0.327	.373	-0.430	.360	-0.470	.326
π_{j0}	0.243*	.144	-1.192	.637	0.348*	.210	0.411*	.203
π_{j1}	0.186**	.139	0.457	.453	-0.017	.215	0.052	.171
π_{j2}	—	—	—	—	0.693*	.399	-0.075	.099
π_{j3}	—	—	—	—	—	—	0.772*	.378

*rejected at 5%
**rejected at 10%

Table 9: Summary statistics for wage equation

	All		Employed		Unemployed	
Observations	3620	—	3217	—	403	—
Age	37.3	(13.1)	37.7	(12.9)	34.3	(14.2)
Higher education	0.30	—	0.31	—	0.25	—
Lower education	0.42	—	0.43	—	0.31	—
Married	0.71	—	0.73	—	0.53	—
Number of children	0.56	(0.74)	0.55	(0.73)	0.62	(0.83)
Local unemployment rate (in %)	7.8	(1.6)	7.9	(1.6)	7.5	(1.5)
Experience	20.6	(14.1)	20.9	(13.8)	18.3	(15.5)

*Standard deviations in parentheses

Table 10: Estimates of the participation and wage offer equation

	Participation		Mean log wage	
Variable	Coef.	p-value	Coef.	p-value
Constant	0.473	.004	4.682	.000
Higher education	0.162	.172	0.303	.000
Lower education	0.409	.001	0.118	.000
Married	0.277	.012	—	—
Number of children	-0.166	.005	—	—
log(Experience)	0.561	.000	0.668	.000
log(Experience)^2	-0.067	.068	-0.146	.000
log(Experience)×log(Age)	—	—	0.353	.000
[log(Experience)×log(Age)]^2	—	—	-0.184	.000
Local unemployment rate	—	—	-0.019	.000
Sigma	—	—	0.424	.000
Rho	—	—	-0.540	.000
No. of observations	2786		2658	
	Log lik.=-1896.53			