Measurement Error Bias Reduction in Unemployment Durations

Montezuma Dumangane*

ISEG-UTL and CEMAPRE

February 2007

Abstract

This paper investigates the impact of duration response measurement error using small parameter asymptotics. The probability limit of GMM type estimators that ignore its presence is derived and illustrated for single spell models with right censored observations, and two a spells lagged duration dependence model. The results suggest an easy-to-estimate adjusted GMM estimator that does not require specification of the measurement error distribution. Identification is achieved by using the moment condition that defines the specification score test sensitive to measurement error. The results are applied to modelling unemployment durations from the BHPS, allowing for heteroskedastic measurement error related to a measure of recall effort.

*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 study was supported in part by Fundação para a Ciência e Tecnologia, program POCI, partially funded by FEDER.

INTRODUCTION

The issue of measurement error has become an increasingly important issue for researchers in all areas of econometrics and in particular outside the linear regression model. The problem arises because the mapping between the theoretical constructs and the observational data is not perfect. Many collected duration data suffer from this limitation. In particular, when event histories are constructed from recall data, the distribution of observed durations often differs from the distribution of the true durations due to contamination with measurement error. This error may assume several forms of which the most common are rounding, heaping or just continuous measurement error. In any case, it leads to a distortion in the properties of duration distributions, for example typically having higher variance, differently shaped density and distorted hazard duration dependence.

This paper considers problems in which measurement error and error free durations are independently distributed. This may occur when spell lenght is collected from information on entry and exit dates from a state, such as unemployment. The aim of the paper is twofold: First to study and understand how this contamination distorts duration distributions. In particular it is investigated how measurement error changes the form of duration dependence in the hazard function and, how this form of duration dependence determines the extent of the impact of the measurement error. The relevance of this analysis comes from the fact that economic theory or previous studies may be informative on the form of the hazard function duration dependence, and that knowledge can be used by practitioners to assess if duration measurement error may or not be a serious threat. Secondly, this paper provides an easy to implement estimator that accounts for the measurement error in duration data, without having to specify its distribution, and applicable when no multiple observations on the same spell are available (often needed to make inference on some features of the distribution of the errors). The importance of developing semiparametric estimators is, in the similar context of uncontrolled heterogeneity, 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.

Few attempts have been made to develop statistical procedures concerned with correcting for this type of 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 fully known parametric model for the measurement error distribution and study its bias properties using small variance approximations. These contributions taken individually suffer from limitations that a practiotioner often finds, namely, the need for multiple observations, the unavaibility of estimates for the shape of the hazard function and the need to specify the error distribution for which economic theory is uninformative.

The approach developed here aims at facing all these difficulties and is used in Chesher, Dumangane and Smith (2002) as means to construct a specification test sensitive to measurement error. This is achieved by considering an approximation to the error contaminated model that incorporates into a known error free distribution measurement error with an unspecified distribution.

In this paper these approximations are used in two ways; to perform specification analysis and to derive a measurement error adjusted GMM estimator. In the first case the effect of duration response measurement error on the probability limit of GMM-estimators constructed ignoring its presence is investigated and illustrated. 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 with correlated errors. The second contribution of the paper suggests a measurement error adjusted GMM estimator of the parameters of the error free distribution and variance of the measurement error, similar to Chesher (2000) for covariate measurement error. The general idea 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 correct the bias in the original moment conditions up to a term that is asymptotically negligible without specifying the distribution of the errors[1].

This approach relates to the correcting estimation equations literature used in the context of covariate measurement error, where the errors are assumed normally distributed. Examples of this approach are the conditional score method of Stefanski and Carroll (1987) applied to generalized linear models (see McCullagh and Nelder, 1989), the method of corrected score equations of Stefanski (1989) and Nakamura (1990) extended in Buzas and Stefanski (1995) for certain generalized linear models, and Bounaccorsi (1996).

[1]The asymptotics here is with respect to the variance of the measurement error.

An important feature of the estimation procedure proposed here is that it 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 score type test sensitive to measurement error in Chesher, Dumangane and Smith (2002) and Dumangane's (2000) extension for multiple-spells-single-destination (MSSD) models. 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.

In the presence of measurement error, the estimator proposed here is approximately consistent, in the sense that the difference between its probability limit and the true parameter is only (sigma) asymptotically zero. When compared with the MLE that ignores measurement error, it is shown that the measurement error adjusted estimator reduces the sigma order of the bias as confirmed by Monte Carlo (see Dumangane, 2006). Examples of other estimators with similar properties are: Chesher (1998, 1999), Wolter and Fuller (1982) and Carroll and Stefanski (1990) for the error-in-variables linear regression, Chesher and Santos Silva (2002) for the heterogeneity adjusted logit model, and Skinner and Humphreys (1999) Weibull model with duration measurement error when the proposed distribution is misspecified. The main pourpose of this estimator is to give practioners that have no information on the distribution of the error (as auxiliary data), and are not willing to make parametric assumptions, a quick and easy way of assessing the impact of duration measurement error on parameter estimates. Ofecourse if the amount of measurement error is not excessive the approximate model can itself be adequate.

As pointed out in Chesher, Dumangane and Smith (2002), when the true duration distribution belongs to the scale-parameter family, 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, which assumes a Gamma distributed random term in the scale parameter, and to Heckman and Singer (1984) and Honoré (1990) estimators for proportional hazards with unspecified unobserved heterogeneity, which assume a known parametric form for the baseline hazard.

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 mean to provide 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?[2]. The identification issue is illustrated in the application where two parametric specifications to model unemployment durations from the BHPS survey are considered. In this case it is shown that misspecifying the distribution of the error free duration produces a conflict in the measurement error corrected GMM estimates of the parameters of error free model and the variance of the measurement error. To take into account that duration measurement error may depend on the recall effort individuals make, inducing some form of heteroskedasticity, its variance is made to dependent on an observed measure of recall effort. This specification also permits to test to what extent the excessive scale variantion can indeed be attributed to measurement error. The results show how biased the analysis can be if the errors are not taken into account and therefore the usefulness of this estimator.

[2]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.

The remainder of 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. Section 3 derives the probability limit of the inconsistent estimator and presents some examples. Section 4 derives the measurement error adjusted GMM estimator. Section 5 applies the estimator to mismeasured unemployment duration from the BHPS survey. Section 6 concludes.

THE EFFECT OF MEASUREMENT ERROR

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 $f_V(v)$[3].

Under this conditions, Chesher, Dumangane and Smith (2002) demonstrate that the small parameter asymptotic approximations for the density and survival functions of S are[4]

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

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

where the second results from integration of (1). Here and later "$\simeq$" denotes an

[3]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.

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

approximation error of order $o(\sigma^2)$ where $\lim_{\sigma\to 0}\frac{o(\sigma^2)}{\sigma^2} = 0$. Sufficient conditions for these approximations to hold are that (a) V has bounded third absolute moment and (b) T is continuously distributed with a density function which has finite and uniformly bounded derivatives up to order 3. This approximations show the local effect on the distribution of duration of this form of measurement error while not depending on the form of the distribution of the errors. In Dumangane (2006) these are used to characterize the impact on the form of duration dependence on the hazard function in a generic setting.

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 can be 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, possibly 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 $\log S_j = \log T_j + \sigma_j U_j$ [5]. This error model is not valid when $\sum_{k=1}^{R} T_k = \sum_{k=1}^{R} S_k$,

[5] 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

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[6]. Under this conditions the density of $\mathbf{S}$ is the $R-$folded integral

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

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 upon collecting terms using the assumptions made on $\mathbf{U}$ to obtain,

$$\begin{aligned} f_{\mathbf{S}}(s) \;\; \simeq \;\; & f_{\mathbf{T}}(\mathbf{t}) + \left\{ \left(\tfrac{1}{2} \sum_{k=1}^{R} \sigma_k^2 + \sum_{k=1}^{R-1} \sum_{l=k+1}^{R} \sigma_{kl} \right) f_{\mathbf{T}}(\mathbf{s}) + \tfrac{3}{2} \sum_{k=1}^{R} \sigma_k^2 s_k f_{\mathbf{T}}^{(k)}\left(\mathbf{s}\right) + \right. \\ & \left. + \sum_{k=1}^{R} \sum_{l \neq k}^{R} \sigma_{kl} s_k f_{\mathbf{T}}^{(k)}(\mathbf{s}) + \tfrac{1}{2} \sum_{k=1}^{R} \sigma_k^2 s_k^2 f_{\mathbf{T}}^{(kk)}(\mathbf{s}) + \sum_{k=1}^{R-1} \sum_{l=k+1}^{R} \sigma_{kl} s_k s_l f_{\mathbf{T}}^{(kl)}(\mathbf{s}) \right\} \end{aligned} \tag{5}$$

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

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 and does not depend on the form of the measurement error joint distribution.

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 of 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

stages of the process might have happen in different points in time demanding a different recall effort for each individual.

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

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

$$E_T[g\,(T, \phi_0)\,|\phi = \phi_0] = 0, \qquad g\,(t, \phi) = d \cdot g_1(t, \phi) + (1 - d) \cdot g_0(t, \phi) \tag{6}$$

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[7]. What follows is valid when all observations are uncensored, by letting c_i go to infinity.

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

Let the density of S_i^* be $f_S(s, \theta_0)$ where 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 (6) are no longer satisfied[8].

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

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

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

$$\arg\max_{\phi} \hat{Q}_n(\phi) = -\hat{g}_n(\phi)' \hat{W} \hat{g}_n(\phi) \tag{7}$$

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

$$E_S[g(S, \widetilde{\phi}(\theta_0)|\theta = \theta_0] = 0 \tag{8}$$

If $\sigma_0^2 = 0$ this gives us (6) 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 $\widetilde{\phi}(\theta_0)$ correct up to a $O(\sigma^3)$ term can be constructed by first order Taylor series expansion around $\sigma_0^2 = 0$. First write (8) in the integral form

$$\int_0^z g_1(s, \widetilde{\phi}(\theta_0)) f_S(s,\theta_0) ds + g_0(z, \widetilde{\phi}(\theta_0)) \bar{F}_S(z,\theta_0) = 0 \tag{9}$$

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 $O(\sigma^2)$ 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 (9) 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

$$b^a(\theta) = \int_0^z g_1(s,\phi) m_T(s,\theta) ds + g_0(z,\phi) M_T(z,\theta) \tag{10}$$

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

$$\widetilde{\phi}(\theta_0) \simeq \phi_0 - (G_0' W G_0)^{-1} G_0' W \; b^a(\theta_0) \tag{11}$$

Expression (11) shows that, up to a term of order $O(\sigma^3)$, the probability limit of $\hat{\phi}_n$ is a linear combination of the bias in the moment conditions induced by measurement error. This is not surprising since GMM estimators have an influence function representation. The usual particular cases apply here, namely (i) when $q = p$, the matrix G_0 is square and expression (11) 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

$$\sqrt{n}(\hat{\phi}_n - \tilde{\phi}(\theta_0)) = N\left[0, (\widetilde{G}_\theta' W \widetilde{G}_\theta)^{-1} \widetilde{G}_\theta' W \ \tilde{\Omega}_\theta \ W' \widetilde{G}_\theta (\widetilde{G}_\theta' W \widetilde{G}_\theta)^{-1}\right] + o_p(1) \qquad (12)$$

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 $\widetilde{G}_\theta \equiv G_\theta(\tilde{\phi}(\theta_0))$.

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

$$\tilde{\boldsymbol{\phi}}(\boldsymbol{\theta}_0) = \boldsymbol{\phi}_0 + \sum_{k=1}^{R} \left.\frac{d\boldsymbol{\phi}}{d\sigma_k^2}\right|_{\boldsymbol{\sigma}=\mathbf{0}} \sigma_k^2 + \sum_{k=1}^{R-1} \sum_{l=k+1}^{R} \left.\frac{d\boldsymbol{\phi}}{d\sigma_{kl}}\right|_{\boldsymbol{\sigma}=\mathbf{0}} \sigma_{kl} + o(||\boldsymbol{\sigma}||) \qquad (13)$$

Let $\tilde{\boldsymbol{\phi}}(\boldsymbol{\theta}_0)$ be the parameter vector that solves the implicit set of equations now defined by $E_{\mathbf{S}}[\mathbf{g}(\mathbf{S}, \tilde{\boldsymbol{\phi}}(\boldsymbol{\theta}_\mathbf{0})) | \boldsymbol{\theta} = \boldsymbol{\theta}_\mathbf{0}] = 0$. Using the approximation to the multiple spell joint density, the $O(||\boldsymbol{\sigma}||)$ probability limit of the inconsistent GMM estimator for this class of models is given by an expression similar to (11), 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

$$\begin{aligned} b^k(\boldsymbol{\phi}_0) &= \int_0^\infty \ldots \int_0^\infty g(\mathbf{s}, \boldsymbol{\phi}_0) \nabla_{\sigma_k^2} f_{\mathbf{S}}^a(\mathbf{s}, \boldsymbol{\phi}_0) ds_R \ldots ds_1 \qquad (14) \\ b^{kl}(\boldsymbol{\phi}_0) &= \int_0^\infty \ldots \int_0^\infty g(\mathbf{s}, \boldsymbol{\phi}_0) \nabla_{\sigma_{kl}} f_{\mathbf{S}}^a(\mathbf{s}, \boldsymbol{\phi}_0) ds_R \ldots ds_1 \end{aligned}$$

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

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,

$$h_T^W(t, x, \alpha, \beta) = \alpha \exp\{\beta' x\} t^{\alpha - 1}, \quad \alpha > 0 \qquad (15)$$

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

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 1 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 (11) for the approximate probability limit does not have a closed form, and therefore numerical integration was needed.

(Figure 1 around here)

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 (11) yields expressions that are up to a term of order $O(\sigma^3)$ equivalent to

$$\begin{aligned} k_W &= \frac{\psi'(1)}{\psi'(1) + \alpha_0^2\sigma_0^2} \\ k_{LL} &= \frac{1 + 2\psi'(1)}{1 + 2\psi'(1) + 3\alpha_0^2\sigma_0^2/4} \end{aligned} \tag{16}$$

where $\psi'(a)$ is the digamma function, and $\psi'(1) = \pi^2/6$. From (16) 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 (16) 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 $\sigma_H^2 = \alpha^2\sigma^2$ - the variance of the random term[9]. A similar result can also be found in Skinner and Humphreys (1999), since there measurement error is implicitly treated as neglected heterogeneity.

[9] 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 (16), it can easily be seen that when there is no censoring the Log-logistic[10] 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.

Example 2 *Two-spell Exponential lagged duration dependence*

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

$$\log \lambda_1 = \gamma_{01} + \gamma_{11}' x, \qquad \log \lambda_2 = \gamma_{02} + \gamma_{12}' x + \delta \log t_1 \tag{17}$$

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,

$$\begin{pmatrix} \tilde{\gamma}_{01}(\boldsymbol{\theta}_0) \\ \tilde{\gamma}_{11}(\boldsymbol{\theta}_0) \\ \tilde{\gamma}_{02}(\boldsymbol{\theta}_0) \\ \tilde{\gamma}_{12}(\boldsymbol{\theta}_0) \\ \tilde{\delta}(\boldsymbol{\theta}_0) \end{pmatrix} \simeq \begin{pmatrix} \gamma_{01} - \frac{\sigma_1^2}{2} \\ \gamma_{11} \\ \gamma_{02} - k_1 - m_1 \frac{\delta_0\sigma_1^2 + \sigma_{12}}{\psi'(1)} \\ \gamma_{12} - \gamma_{11} \frac{\delta_0\sigma_1^2 + \sigma_{12}}{\psi'(1)} \\ \delta_0 - \frac{\delta_0\sigma_1^2 + \sigma_{12}}{\psi'(1)} \end{pmatrix} \tag{18}$$

The following points are of interest:

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

1. Since there is no form of duration dependence in the first spell the regressor coefficients are still consistently estimated[11].

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

$$\frac{\tilde{\delta}(\boldsymbol{\theta}_0)}{\delta_0} = \frac{\psi'(1)}{\psi'(1) + \sigma_1^2 + \sigma_{12}/\delta_0} \tag{19}$$

   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.

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

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 to derive an measurement error adjusted GMM estimator.

BIAS CORRECTED GMM ESTIMATOR

Single spell models

Let the model for T be characterized by the set of moment conditions defined in (6). 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

$$\int_0^z (g_1(s,\phi) - b_1^a(s,\theta))f_S(s,\theta)ds - r_1(z,\theta) + (g_0(z,\phi) - b_0^a(z,\theta))\bar{F}_S(s,\theta) \simeq 0 \quad (20)$$

Because of the order of the approximation considered here, terms of order $O(\sigma^m)$ with $m > 2$ can be omitted. It follows from (10) 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 (10) to

$$\int_0^z b_1^a(s,\theta)f_T(s,\phi)ds + r_1(z,\theta) \quad (21)$$

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

$$\begin{aligned} b_1^a(s,\theta) &= \tfrac{\sigma^2}{2}[sg_1'(s,\phi) + s^2 g_1''(s,\phi)] \quad (22) \\ r_1(z,\theta) &= \tfrac{\sigma^2}{2}\{[zg_1(z,\phi) - z^2 g_1'(z,\phi)]f_T(z,\phi) + z^2 g_1(z,\phi)f_T'(z,\phi)\} \end{aligned}$$

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

$$b^a(s,z,\theta) = d\ b_1^a(s,\theta) + (1-d)\ b_0^a(z,\theta) + r_1(z,\theta) \tag{23}$$

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 (23) that the bias corrected moment conditions are

$$E_S[g^c(S,z,\theta)|\theta=\theta] \simeq 0, \qquad g^c(s,z,\theta) = g(s,z,\phi) - b^a(s,z,\theta) \tag{24}$$

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.

Multiple spell models

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

$$\mathbf{b}^a(\mathbf{s},\theta) = \tfrac{1}{2}tr(\Sigma \text{diag}(\mathbf{s})\text{diag}(\mathbf{G}_{\mathbf{T}}^{(1)})) + \tfrac{1}{2}\iota'\Sigma \otimes (\mathbf{s}'\mathbf{s}) \otimes \mathbf{G}_{\mathbf{T}}^{(2)}\iota \tag{25}$$

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 (23).

Identification and estimation

The measurement adjusted 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 proposed estimator is based on the $(q + 1)$ set of moment conditions

$$E_S[g_e^c(S, z, \theta_0)|\theta = \theta_0] \simeq 0 \tag{26}$$

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 (see for example Newey and McFadden 1994), the proposed GMM estimator $\hat{\theta}_n^c$, is defined as $\arg\max \hat{Q}_{e,n}^c(\theta) = -\hat{g}_{e,n}^c(\theta)'\hat{g}_{e,n}^c$, with first order asymptotic distribution

$$\sqrt{n}(\hat{\theta}_n^c - \theta^a) = N\left[0, (G_e^a)^{-1} \; \Omega_e^a \; (G_e^a)^{-1\prime}\right] + o_p(1) \tag{27}$$

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 $O(\sigma^3)$ 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 (2) and (3) around here)

Figures (2) and (3) 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[12].

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 varies 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 $b_{\boldsymbol{\sigma}}^a(s, z, \theta)$ denotes the approximate structural bias function associated with

[12] The performance of this estimator is investigated via Monte Carlo in Dumangane (2006).

$D_{\sigma^2}(t,c,\phi)$, then estimation of $\pi' = (\pi_0, \pi_1')$ requires the additional estimating equations:

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

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.

ERROR CONTAMINATED UNEMPLOYMENT DURATIONS

Several studies using the British Household Panel Survey (BHPS) have examined recall error in unemployment durations, see for example Paull (1996), Elias (1996), Dex and McCulloch (1997) and Brendan (1997). In this application data from the BHPS collected at wave 1 is used. 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. Thus the potential for error contaminated duration is high[13].

[13] 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 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.

The model

Interest lies on estimation of the parameters of the reduced form conditional hazard function of time to leaving unemployment. Early examples of this approach are Lancaster (1979) and Nickell (1979 a,b) whereas Narendranathan, Nickell and Stern (1985) provide an excellent discussion on the effect of unemployment benefits in unemployment duration. This later work was the basis for choosing the economic specification. The conditional distribution of time to leaving unemployment will be a function of the following set of exogenous variables in table 1

(Table 1 around here)

Table 2 shows descriptive statistics of the data before being transformed. As expected, those who never had any form of unemployment benefit are on average younger, more educated and experienced smaller spells of unemployment.

(Table 2 around here)

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, two distinct populations are sampled: those belonging to the

---

the reported month.

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. From both, at time T_I the i-ith individual date of interview, a complete or censored duration is recorded from retrospective information on the entry and exit dates from unemployment, T_E and T_X respectively.

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,

$$f_T^*(t, x, \phi) = \frac{t f_T(t, x, \phi) + E[\triangle T] f_T(t, x, \phi)}{E[T] + E[\triangle T]} \tag{28}$$

It is a weighted average of the likelihood of a stock sample and a flow sample[14]. 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 sub-population. 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 (28) 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 concave hazard functions, representing an unemployment process in which the risk of leaving unemployment reaches a peak, and becomes persistent as time goes by. Since one of these models has to be misspecified the aim is to see how the GMM estimator behaves in such a case.

First the maximum likelihood estimates are presented together with a specification analysis.

[14] I am in debt to Andrew Chesher for deriving this sampling distribution.

Maximum likelihood estimates

Table 3 reports maximum likelihood estimates ignoring the presence of measurement error and its adjusted GMM counterparts for both specifications[15].

(Table 3 around here)

The MLE estimate of the Weibull shape parameter suggests a decreasing hazard for leaving unemployment and the constant hazard rate model is rejected. The Log-logistic MLE estimate identifies a non-monotonic hazard, being the hypothesis of a monotonic decreasing hazard rate rejected. An empirical researcher's next step would be to perform specification analysis of both models. Two procedures were used, a general one and one specific to detect duration response measurement error. Firstly, residual analyses as a means to investigate the general quality of the fit of the MLE estimates is performed. The principle is that under the null hypothesis of homogeneity, the integrated hazard vector are n realizations of a mean-one Exponential variate (see Lancaster and Chesher, 1985).

(Figure 4 around here)

It is clear from Figure 4 that 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, even if error contaminated, than the Weibull.

Secondly, the eficient version of the measurement error sensitive specification test in Chesher, Dumangane and Smith (2002) is reported in the bottom of Table (3)[16].

[15] 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).

[16] 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

At a 0.05 nominal level, the null hypothesis is clearly rejected in the Weibull model, suggesting the presence of measurement error or, indeed other misspecifications as incorrect hazard specification or uncontrolled heterogeneity, but it is not rejected for the Log-logistic model[17].

The measurement error adjusted estimates

The Weibull GMM estimates illustrate what happens if the error free distribution is uncorrectly specified. Despite the estimate of α being now bigger, as it should be if measurement error was present, the estimate of the variance of measurement error is not consistent with it. Clearly this contradiction arises because in this case the unknown probability limit of the GMM estimator will surely depend on how far the assumed error free distribution is from the true. If the hazard is indeed nonmonotonic, as suggested by the Log-logistic, such functional form that can never be captured by a Weibull specification. This is a favourable case in the sense that hazard misspecification can not be captured by response error.

The specification analysis shows that the data is well approximated by the Log-logistic model, nevertheless the need for the GMM estimator in this data set can be easily motivated by the presence of duration measurement error. Recall that this procedure adjusts for any type of misspecification that induces excessive scale variation. That is clearly an advantage for an applied researcher whose interest are the parameters of the error free duration distribution, but imposes an obvious limitation in the interpretation of the parameter σ^2. Bearing this in mind, the results interpretation will take that parameter to be the variance of measurement error, assumption that

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.

[17] This may indicate correct specification or lack of power of the measurement error specification for the characteristics of this data (sample size and censoring proportion).

will be tested in the next section by making it a function of a measurement error related variable

The adjusted GMM estimate of α is consistent with the effect and amount of measurement error in this parameter as predicted in the analysis of section 3:

1. The correction has the right sign, and according to the estimate σ^2 15% of the observed variation in the log durations is attributed to response error.

2. For these estimate of σ^2 and α, 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 predicted, equal to correction in the shape parameter. The observed proportional bias for the statistically significant coefficients varies from 0.90 to 0.95.

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 three comments: First, 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. Secondly, the higher the level of income in unemployment the lower the risk of exit[18].

[18] 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.

Thirdly 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, perhaps because the rate of job offers is a negative function of the mean wage.[19].

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

(Figure 5 around here)

Figure 5 shows the estimated hazard functions for the two set of estimates, conditional on two cohorts characterized by whether the individual receives any type of income support. The covariates are evaluated at the sample means of each cohort.

As predicted by the approximations (see Dumangane 2006), the correction initially raises the hazard function above the MLE hazard such that the duration at which the risk of exiting reaches its maximum is now smaller[20]. Since from the point of view of efficacy, unemployment policies should target individuals on the increasing part of the hazard, policies based on error contaminated durations are sub-optimal. Note that this issue is particularly relevant for beneficiaries, where the duration at which the MLE hazard reaches its maximum is nearly 50% larger.

[19] 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 this 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.

[20] 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.

These results assume that measurement error is homogenous. The next section estimates the Log-logistic model allowing for heteroskedastic measurement error, reflecting the fact that spells occurring in different time periods being likely to be contaminated with different amounts of measurement error.

Heteroskedastic measurement error

The variance of the measurement error distribution is now defined to be a function of the recall effort individuals undergo. With this specification it is possible to test if the excessive scale parameter variation is, at least partially, indeed due to measurement error or if it is just a consequence of functional form misspecification or uncontrolled heterogeneity.

Since durations are constructed from the entry and exit dates in the state, an appropriated measure of the recall effort w, is the logarithm of 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. It is independent of the spell lenght (required for the validity of the approximation results), 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.

Four specifications for the skedastic function are considered: the linear specification, which can be thought of as a first order local approximation to the true; the exponential specification, which is always a natural candidate for skedastic functions; and 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 4 shows the results for these specifications for the Log-logistic model[21]. The only specification that clearly rejects the skedastic function is the exponential. Its

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

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 4 around 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, but on the other hand the estimates are less precise. 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. After this quantile a 1% variation on the recall effort induces an increase of the measurement error variance of 0,007. This suggests that at least after some point the excessive scale variation is indeed measurement related.

CONCLUSION

The impact of duration response measurement error on parameter estimates was studied by deriving the probability limit of GMM estimators. For single spell models, generally measurement error dampens the form of duration dependence of 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 co-variates affect the duration distribution, in the same fashion as it does for the shape parameters. It was shown that allowing for right censoring has different implications in different parametric specifications. In the Weibull model, right censoring offsets in an increasing way the impact of measurement error in the probability limit of the MLE, while in the Log-logistic it 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.

The measurement error adjusted GMM estimator corrects (approximately) the bias in the moment conditions that define the error-free model. Whenever the error free distribution is known, the moment condition that defines the measurement error specification score test, was shown to provide valuable information about the true parameters. As such under this estimating procedure score tests can be constructive. The main advantage of this estimator is that it does not require any prior information on the measurement error distribution, whose parameters are estimated jointly with the parameters of the error free distribution. Given the small sigma nature of the approximations, when there is evidence of a large amount of measurement error it is worthile considering a procedure that specifies parametrically the distribution of the error. In this case the adjusted estimator can give an idea of the need to find such alternative procedures. If the contamination is only small, estimates obtained from this procedure may be adequate.

Though motivated by measurement error this estimator may identify a non-zero variance if other forms of misspecification are present in the data. Namely, uncontrolled heterogeneity whenever the error free distribution belongs to the scale parameter family, or if the error free model is misspecified. In the last case the estimate of the variance will not have a structural interpretation, but it can in some cases still be useful, as it can be interpreted as the cost of choosing a parametric model for the durations. Note that in the application it was shown that misspecifying the error free distribution does not always produce a non-zero estimate of the variance.

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 MLE 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. Clearly this is a case when interpretation of the variance may be dubious since the model is expected to be misspecified because of uncontrolled heterogeneity or even functional form misspecification. Because of that the variance of the measurement error was made to depend on a measure of recall effort. This specific form of heteroskedasticity is specific to this specification problem and allows to conclude that in this data recall error is indeed present. Distinguishing duration response measurement form uncontrolled heterogeneity may be possible if multiple observations on the same spell are available, but for the applied researcher what really matters is to known the parameters of the error free distribution. If that is the pourpose the GMM estimator provided is a quick and unexpensive way of assessing excessive scale variation.

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[22].

[22] 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

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 5 around 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.

Table 5 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 6 shows the results for both equations.

(Table 6 around here)

Non statistically significant variables were deleted from the equations as it was to be used for prediction.

---

earnings, this variable is closely correlated with the reservation wage and with on job earnings.

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

$$\int_0^z g_1(s,\phi)m_T(s,\theta)ds = \int_0^z b_1^a(s,\theta)f_T(s,\phi)ds + r_1(z,\theta) \tag{29}$$

Using the definition of $m_T(s,\theta)$ the left hand side of (29) can be written as

$$\tfrac{\sigma^2}{2}\left(\int_0^z g_1(s,\phi)f_T(s,\phi)\,ds + 3\int_0^z g_1(s,\phi)sf_T'(s,\phi)\,ds + \int_0^z g_1(s,\phi)s^2 f_T''(s,\phi)\,ds\right) \tag{30}$$

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

$$\begin{array}{llll} A.B.1 & \lim_{s_k\to 0} g_1(s,\phi)sf_T(s,\phi) & = & 0 \\ A.B.2 & \lim_{s_k\to 0} g_1(s,\phi)s^2 f_T'(s,\phi) & = & 0 \\ A.B.3 & \lim_{s_k\to 0} g_1'(s,\phi)s^2 f_T(s,\phi) & = & 0 \end{array} \tag{31}$$

leads to the desired result.

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 single spell single destination case when $R = 1$.

Computation of $E_{\mathbf{S}}[\mathbf{g}(\mathbf{S},\boldsymbol{\phi})|\boldsymbol{\theta}=\boldsymbol{\theta}]$, up to $o(||\boldsymbol{\sigma}||)$ requires the approximation (5) to the multiple spell joint density leadint to calculation of three integrals,

$$
\begin{array}{lll}
1. & \int_0^\infty \ldots \int_0^\infty \mathbf{g}\left(\mathbf{s},\boldsymbol{\phi}\right) s_k f_{\mathbf{T}}^{(k)}\left(\mathbf{s}\right) ds_R \ldots ds_1 & = -(E_{\boldsymbol{\phi}}[\mathbf{g}(\mathbf{S},\boldsymbol{\phi})] + E_{\boldsymbol{\phi}}[S_k \mathbf{g}^{(k)}(\mathbf{S},\boldsymbol{\phi})]) \\
2. & \int_0^\infty \ldots \int_0^\infty \mathbf{g}\left(\mathbf{s},\boldsymbol{\phi}\right) s_k^2 f_{\mathbf{T}}^{(kk)}\left(\mathbf{s}\right) ds_R \ldots ds_1 & = 2E_{\boldsymbol{\phi}}[\mathbf{g}(\mathbf{S},\boldsymbol{\phi})] + 4E_{\boldsymbol{\phi}}[S_k \mathbf{g}^{(k)}(\mathbf{S},\boldsymbol{\phi})] + \\
 & & + E_{\boldsymbol{\phi}}[S_k^2 \mathbf{g}^{(kk)}(\mathbf{S},\boldsymbol{\phi})] \\
3. & \int_0^\infty \ldots \int_0^\infty \mathbf{g}\left(\mathbf{s},\boldsymbol{\phi}\right) s_k s_l f_{\mathbf{T}}^{(kl)}\left(\mathbf{s}\right) ds_R \ldots ds_1 & = -(E_{\boldsymbol{\phi}}[\mathbf{g}(\mathbf{S},\boldsymbol{\phi})] + E_{\boldsymbol{\phi}}[S_k \mathbf{g}^{(k)}(\mathbf{S},\boldsymbol{\phi})] + \\
 & & + E_{\boldsymbol{\phi}}[S_l \mathbf{g}^{(l)}(\mathbf{S},\boldsymbol{\phi})] + E_{\boldsymbol{\phi}}[S_k S_l \mathbf{g}^{(kl)}(\mathbf{S},\boldsymbol{\phi})])
\end{array}
\tag{32}
$$

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,

$$
\begin{array}{llclcl}
A.C.1 & \lim_{s_k \to 0} \mathbf{g}(\mathbf{s},\boldsymbol{\phi}) s_k f_{\mathbf{T}}\left(\mathbf{s}\right) & = & \lim_{s_k \to \infty} \mathbf{g}(\mathbf{s},\boldsymbol{\phi}) s_k f_{\mathbf{T}}\left(\mathbf{s}\right) & = & 0 \\
A.C.2 & \lim_{s_k \to 0} \mathbf{g}(\mathbf{s},\boldsymbol{\phi}) s_k^2 f_{\mathbf{T}}^{(k)}\left(\mathbf{s}\right) & = & \lim_{s_k \to \infty} \mathbf{g}(\mathbf{s},\boldsymbol{\phi}) s_k^2 f_{\mathbf{T}}^{(k)}\left(\mathbf{s}\right) & = & 0 \\
A.C.3 & \lim_{s_k \to 0} \mathbf{g}(\mathbf{s},\boldsymbol{\phi}) s_k s_l f_{\mathbf{T}}^{(l)}\left(\mathbf{s}\right) & = & \lim_{s_k \to \infty} \mathbf{g}(\mathbf{s},\boldsymbol{\phi}) s_k s_l f_{\mathbf{T}}^{(l)}\left(\mathbf{s}\right) & = & 0 \\
A.C.4 & \lim_{s_l \to 0} \mathbf{g}^{(k)}(\mathbf{s},\boldsymbol{\phi}) s_k s_l f_{\mathbf{T}}\left(\mathbf{s}\right) & = & \lim_{s_l \to \infty} \mathbf{g}^{(k)}(\mathbf{s},\boldsymbol{\phi}) s_k s_l f_{\mathbf{T}}\left(\mathbf{s}\right) & = & 0
\end{array}
\tag{33}
$$

The approximate required expectation can now be written as

$$
E_{\mathbf{S}}[\mathbf{g}(\mathbf{S},\boldsymbol{\phi})|\boldsymbol{\theta}=\boldsymbol{\theta}] \simeq a_1(\boldsymbol{\sigma}) E_{\mathbf{T}}[\mathbf{g}(\mathbf{S},\boldsymbol{\phi})|\boldsymbol{\phi}=\boldsymbol{\phi}] + a_2(\boldsymbol{\sigma}) E_{\mathbf{T}}[S_k \mathbf{g}^{(k)}(\mathbf{S},\boldsymbol{\phi})|\boldsymbol{\phi}=\boldsymbol{\phi}]
$$

$$
+ a_3(\boldsymbol{\sigma}) E_{\mathbf{T}}[S_k^2 \mathbf{g}^{(kk)}(\mathbf{S},\boldsymbol{\phi})|\boldsymbol{\phi}=\boldsymbol{\phi}] + a_4(\boldsymbol{\sigma}) E_{\mathbf{T}}[S_k S_l \mathbf{g}^{(kl)}(\mathbf{S},\boldsymbol{\phi})|\boldsymbol{\phi}=\boldsymbol{\phi}] \tag{34}
$$

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

$$
\begin{aligned}
a_1(\boldsymbol{\sigma}) &= -\frac{3}{2}\sum_{k=1}^{R}\sigma_k^2 - \sum_{k=1}^{R}\sum_{l\neq k}^{R}\sigma_{kl} + \sum_{k=1}^{R}\sigma_k^2 + \frac{1}{2}\sum_{k=1}^{R}\sigma_k^2 + 2\sum_{k=1}^{R-1}\sum_{l=k+1}^{R}\sigma_{kl} \qquad (35) \\
&= -\sum_{k=1}^{R}\sum_{l\neq k}^{R}\sigma_{kl} + 2\sum_{k=1}^{R-1}\sum_{l=k+1}^{R}\sigma_{kl} = 0
\end{aligned}
$$

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

$$\begin{aligned} a_2(\boldsymbol{\sigma}) &= -\frac{3}{2}\sum_{k=1}^{R}\sigma_k^2 \bar{w}_k(\boldsymbol{\phi}) - \sum_{k=1}^{R}\sum_{l\neq k}^{R}\sigma_{kl}\bar{w}_k(\boldsymbol{\phi}) + 2\sum_{k=1}^{R}\sigma_k^2 \bar{w}_k(\boldsymbol{\phi}) + \\ &\quad + \sum_{k=1}^{R-1}\sum_{l=k+1}^{R}\sigma_{kl}\bar{w}_k(\boldsymbol{\phi}) + \sum_{k=1}^{R-1}\sum_{l=k+1}^{R}\sigma_{kl}\bar{w}_l(\boldsymbol{\phi}) \end{aligned} \tag{36}$$

Noting that

$$\sum_{k=1}^{R}\sum_{l\neq k}^{R}\sigma_{kl}\bar{w}_k(\boldsymbol{\phi}) = \sum_{k=1}^{R-1}\sum_{l=k+1}^{R}\sigma_{kl}\bar{w}_k(\boldsymbol{\phi}) + \sum_{k=1}^{R-1}\sum_{l=k+1}^{R}\sigma_{kl}\bar{w}_l(\boldsymbol{\phi}) \tag{37}$$

gives

$$a_2(\boldsymbol{\sigma}) = \frac{1}{2}\sum_{k=1}^{R}\sigma_k^2 \bar{w}_k(\boldsymbol{\phi}) \tag{38}$$

3. Finally the terms associated with

$$E_{\mathbf{T}}[S_k^2 \mathbf{g}^{(kk)}(\mathbf{S}, \boldsymbol{\phi})|\boldsymbol{\phi} = \boldsymbol{\phi}] \equiv \bar{w}_{kk}(\boldsymbol{\phi}); \qquad E_{\mathbf{T}}[S_k S_l \mathbf{g}^{(kl)}(\mathbf{S}, \boldsymbol{\phi})|\boldsymbol{\phi} = \boldsymbol{\phi}] \equiv \bar{w}_{kl}(\boldsymbol{\phi}), \tag{39}$$

are respectively

$$a_3(\boldsymbol{\sigma}) = \frac{1}{2}\sum_{k=1}^{R}\sigma_k^2 \bar{w}_{kk}(\boldsymbol{\phi}); \qquad a_4(\boldsymbol{\sigma}) = \sum_{k=1}^{R-1}\sum_{l=k+1}^{R}\sigma_{kl}\bar{w}_{kl}(\boldsymbol{\phi}) \tag{40}$$

It follows that the approximate structural bias function is given by

$$\mathbf{b}^a(\mathbf{s}, \boldsymbol{\theta}) = \frac{1}{2}\sum_{k=1}^{R}\sigma_k^2\left(s_k \mathbf{g}^{(k)}(\mathbf{s}, \boldsymbol{\phi}) + s_k^2 \mathbf{g}^{(kk)}(\mathbf{s}, \boldsymbol{\phi})\right) + \sum_{k=1}^{R-1}\sum_{l=k+1}^{R}\sigma_{kl}s_k s_l \mathbf{g}^{(kl)}(\mathbf{s}, \boldsymbol{\phi}). \tag{41}$$

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

$$E[g^a(Y, \theta_0)|\theta = \theta_0] = O\left(\sigma_0^3\right) \tag{42}$$

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 3 *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*

$$\theta_0 = \arg\max_{\theta} Q_0(\theta) = -g_0(\theta)'Wg_0(\theta) \tag{43}$$

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

$$\theta^a = \arg\max_{\theta} Q^a_0(\theta) = -g^a_0(\theta)'Wg^a_0(\theta) \tag{44}$$

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

The proof exploits the fact that $\hat{\theta}^a_n$ 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,

$$G_n^a(\hat{\theta}_n^a) W g_n^a(\hat{\theta}_n^a) = 0 \tag{45}$$

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

$$G_n^a(\hat{\theta}_n^a)' W [g_n^a(\theta_0) + G_n^a(\ddot{\theta}_n)(\hat{\theta}_n^a - \theta_0)] = 0 \tag{46}$$

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

$$n^{1/2}(\hat{\theta}_n^a - \theta_0) = -[G_n^a(\hat{\theta}_n^a)' W G_n^a(\ddot{\theta}_n)]^{-1} G_n^a(\hat{\theta}_n^a) W n^{1/2} \; g_n^a(\theta_0) \tag{47}$$

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

$$n^{1/2}(\hat{\theta}_n^a - \theta_0) = A \; n^{1/2}(g_n^a(\theta_0) - O(\sigma_0^3)) + n^{1/2} O(\sigma_0^3) + o_p(1) \tag{48}$$

or equivalently

$$n^{1/2}(\hat{\theta}_n^a - \theta_0 - O(\sigma_0^3)) = A n^{1/2}(g_n^a(\theta_0) - O(\sigma_0^3)) + o_p(1) \tag{49}$$

It follows directly from (42) that $n^{1/2}(g_n^a(\theta_0) - O(\sigma_0^3)) \xrightarrow{d} N(0, V)$ where

$$V = E[(g^a(Y, \theta_0) - O(\sigma_0^3))(g^a(Y, \theta_0) - O(\sigma_0^3))' | \theta = \theta_0] \tag{50}$$

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

$$\hat{\theta}_n^a - \theta_0 \xrightarrow{p} O(\sigma_0^3). \tag{51}$$

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 and Nicoletta Rosati for helpful suggestions. I am grateful 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 research was supported by Fundação para a Ciência e Tecnologia and by PRAXIS XXI.

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

[3] Buzas, J.S., and L.A. Stefanski (1995): "A note on corrected score estimation," *Statistics & Probability Letters*, 28, 1-18.

[4] 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.

[5] Cavanagh, C., and R.P. Sherman (1998): "Rank estimators for Monotonic Index models," *Journal of Econometrics*, 84, 351-381.

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

[7] Chesher, A.D., Dumangane, M., and R.J. Smith (2002): "Duration response measurement error," *Journal of Econometrics*, 111, 169-194.

[8] Chesher, A.D., and J.M.S. Silva (2002): "Taste variation in Discrete Choice models," *Review of Economics Studies*, 69, 147-168.

[9] Cox, D.R. (1972): *Renewal Theory.* London: Metheun.

[10] 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.

[11] Godfrey, L.G. (1988): *Misspecification Tests in Econometrics.* Cambridge: Cambridge University Press.

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

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

[14] 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.

[15] 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.

[16] Honore, B. (1990): "Identification results for duration models with multiple spells or time-varying covariates," unpublished manuscript, Northwestern University.

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

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

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

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

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

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

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

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

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

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

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

[28] 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.

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

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

[31] Pace, L., and A. Salvan (1997): "*Principles of Statistical Inference From a Neo-Fisherian Perspective*" . London: World Scientific.

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

[33] Skinner, C.J., and K. Humphreys (1999) "Weibull regression for lifetimes measured with error," *Lifetime Data Analysis*, 5, 23-37.

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

[35] 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.

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

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

[38] Wolter, K.M., and W.A. Fuller (1982): "Estimation of nonlinear errors-in-variables models," *Annals of Statistics*, 10, 539-548.

Table 1. Definition of covariates.

Covariate	Description
AGE	Logarithm of age.
HEDUC	1 if Higher or first degree, teaching qualification and other qualifications.
LEDUC	1 if CSE, commercial, GCE and nursing qualifications, apprenticeship, and other lower qualifications.
MARR	1 if married or leaving as a couple.
NCH	No. of dependent children in the household.
UNRATE	Unemployment rate at the metropolitan area of residence.
BENEF	Benefeciary
INCUN	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.
INWK	The log of weekly estimated earnings specified as a function of work experience measures and other individual characteristics

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.

Table 2. Summary statistics for wave 1 sample.

	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)
HEDUC	0.23	—	0.22	—	0.30	—
LEDUC	0.42	—	0.41	—	0.47	—
MARR	0.56	—	0.59	—	0.44	—
NCH	0.56	(0.7)	0.56	(0.8)	0.59	(0.7)
UNRATE (in %)	7.8	(1.6)	7.9	(1.6)	7.5	(1.5)
INCUN (£ per week)	53.5	(55.5)	67.6	(54.2)	—	—
INCWK (£ per week)	160.1	(52.7)	161.0	(48.3)	156.9	(66.8)

* Standard errors in parentheses for continuous variables

Table 3. 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
HEDUC	1.322*	.282	1.538*	.270	1.713*	.342	1.798*	.328
LEDUC	0.702*	.284	1.014*	.244	1.119*	.339	1.215*	.334
MARR	0.636*	.199	0.729*	.196	0.862*	.256	0.938*	.264
NCH	-0.372*	.135	-0.476*	.137	-0.506*	.158	-0.564*	.166
UNRATE	-0.101**	.055	-0.086	.058	-0.148*	.070	-0.164*	.075
BENEF	-0.784*	.258	-1.116*	.200	-1.203*	.235	-1.346*	.256
INCUN	-0.423*	.059	-0.431*	.063	-0.612*	.099	-0.653*	.112
INCWK (B)	-0.269	.354	-0.591**	.305	-0.488	.438	-0.482	.446
INCWK (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 4. 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
HEDUC	1.828*	.350	1.880	.415	1.937*	.392	1.939*	.378
LEDUC	1.265*	.361	1.335	.436	1.392*	.397	1.389*	.380
MARR	0.979*	.277	1.015	.327	1.039*	.316	1.031*	.301
NCH	-0.600*	.175	-0.632	.211	-0.660*	.208	-0.654*	.203
UNRATE	-0.177*	.077	-0.186	.088	-0.198*	.085	-0.195*	.082
BENEF	-1.399*	.261	-1.453	.315	-1.481	.317	-1.469*	.297
INCUN	-0.652*	.114	-0.670	.133	-0.685*	.132	-0.681*	.125
INCWK (B)	-0.488	.455	-0.488	.522	-0.560	.503	-0.580	.484
INCWK (N. B.)	-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 5. Summary Statistics for sample

	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 6. Estimates of the probit and wage offer equation

	Participation		Mean log wage	
Variable	Coef.	p-value	Coef.	t-ratio
Constant	0.473	.004	4.682	.000
HEDUC	0.162	.172	0.303	.000
LEDUC	0.409	.001	0.118	.000
MARR	0.277	.012	—	—
NCH	-0.166	.005	—	—
log(EXPER)	0.561	.000	0.668	.000
$\log(\text{EXPER})^2$	-0.067	.068	-0.146	.000
log(EXPER)×log(AGE)	—	—	0.353	.000
$[\log(\text{EPER})\times\log(\text{AGE})]^2$	—	—	-0.184	.000
UNRATE	—	—	-0.019	.000
Sigma	—	—	0.424	.000
Rho	—	—	-0.540	.000
No. of observations	2786		2658	
	Log lik.=-1896.53			

FIG. 1. Approximate proportional inconsistency of α and β_1, as a function of the censoring proportion for $Var(\log T)/Var(\log S) \in \{0.80, 0.85, 0.90\}$.

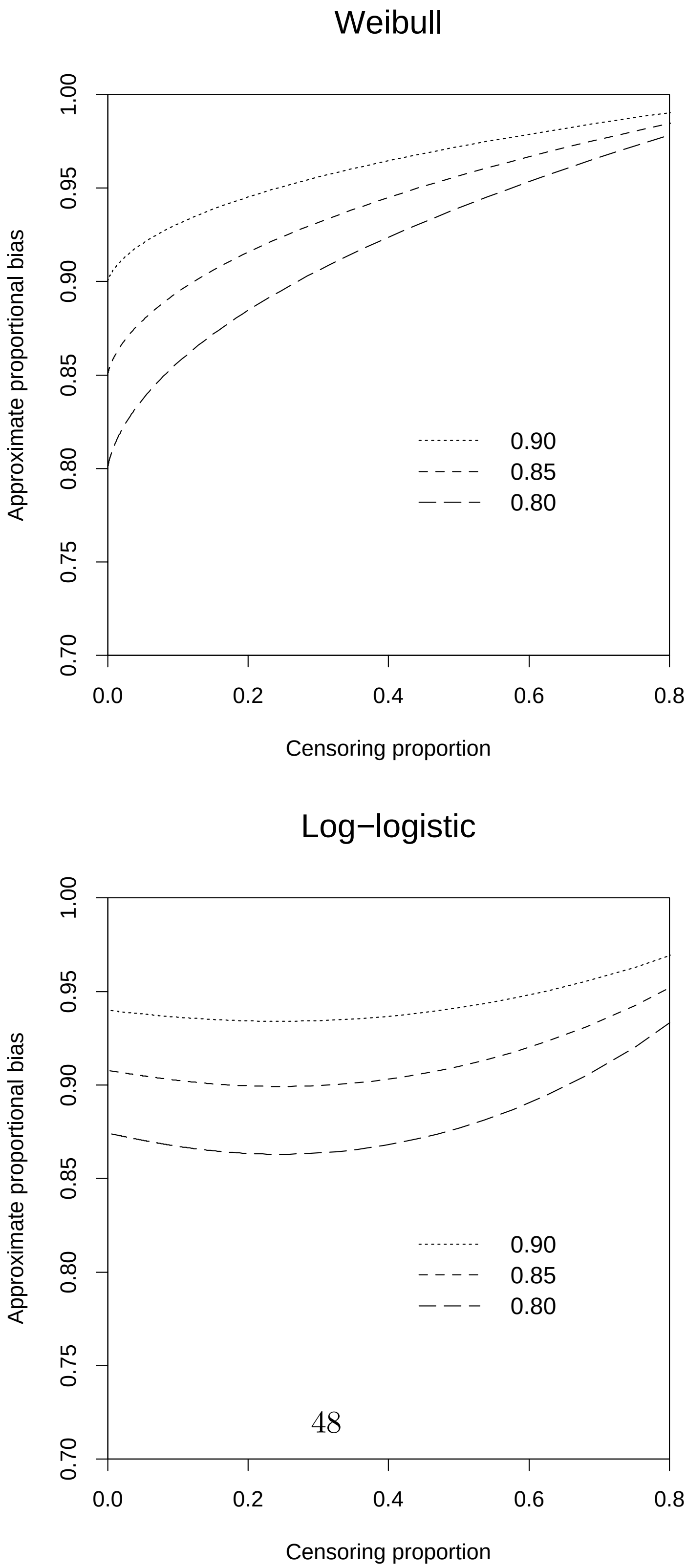

FIG. 2. Exact expectation of Weibull scores and approximate bias corrected scores with Lognormal (dotted) and Gamma (dashed) measurement error for $\alpha \in \{0.8, 1, 1.5\}$ and 20% and 50% censoring.

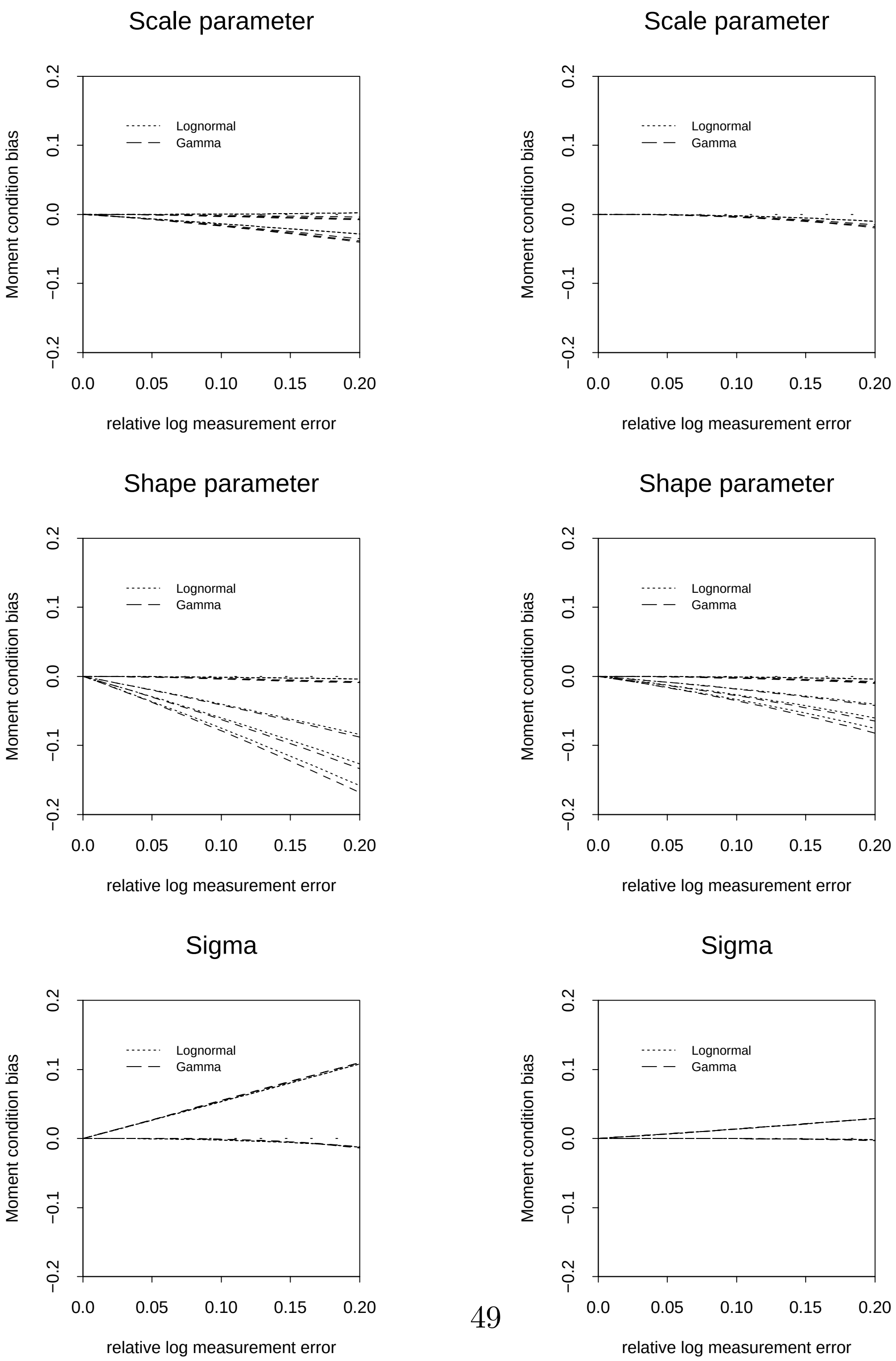

FIG. 3. Exact expectation of Log-logistic scores and approximate bias corrected scores with Lognormal (dotted) and Gamma (dashed) measurement error for $\alpha \in \{0.8, 1, 1.5\}$ and 20% and 50% censoring.

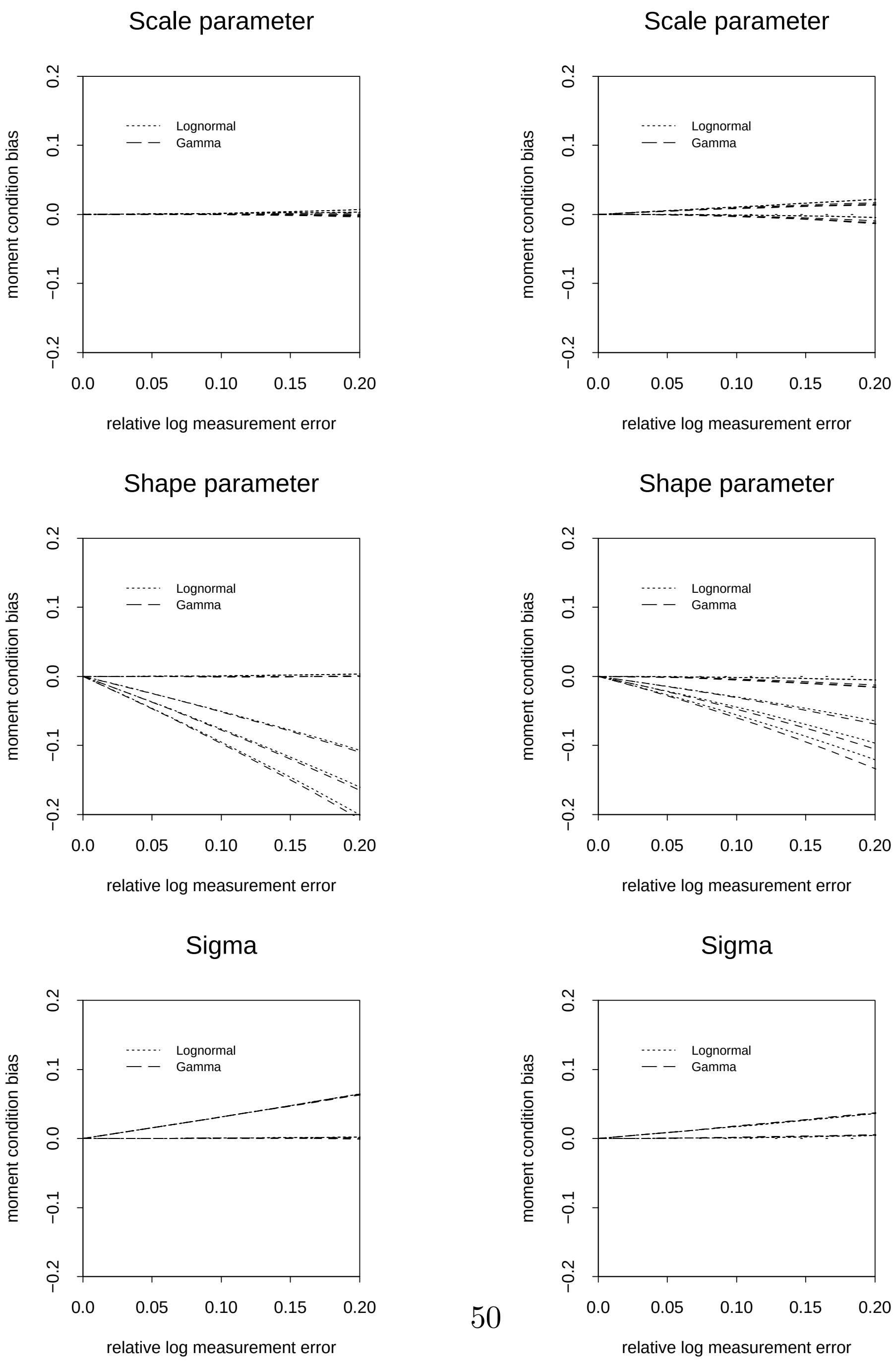

FIG. 4. Residual analyses for Weibull and Log-logistic MLE estimates.

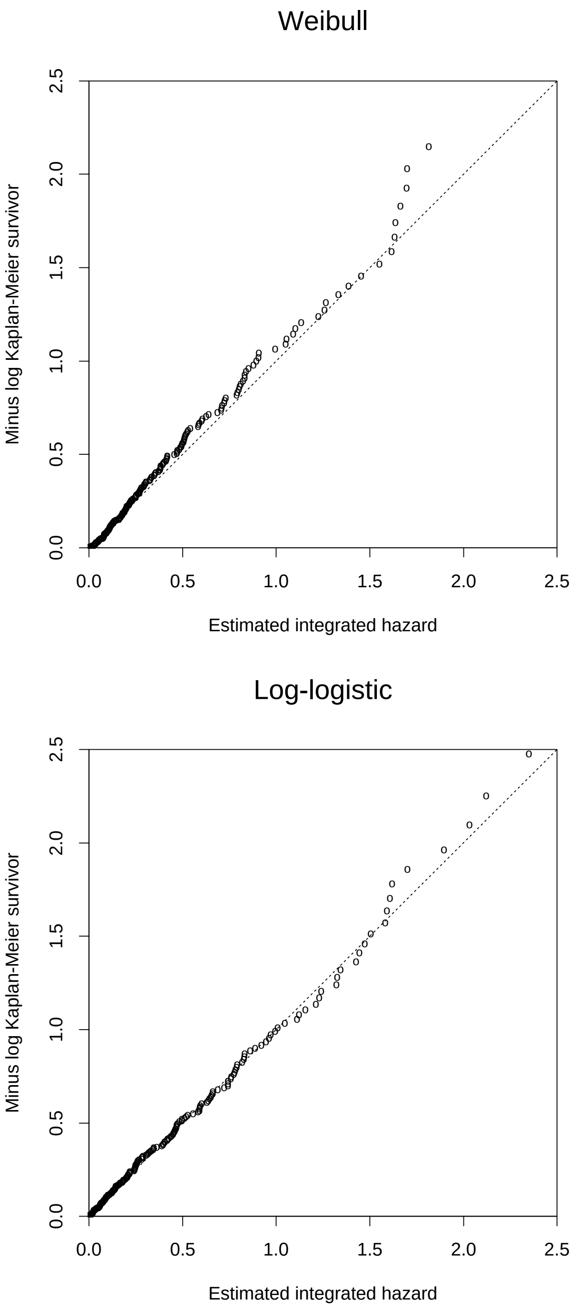

FIG. 5. Estimated error contaminated (MLE) and bias corrected (GMM) hazard functions.