Behavior Research Methods, Instruments, & Computers
2003, 35 (4), 485-492

QMLE: Fast, robust, and efficient estimation of distribution functions based on quantiles

SCOTT BROWN
University of California, Irvine, California

and

ANDREW HEATHCOTE
University of Newcastle, Newcastle, New South Wales, Australia

Quantile maximum likelihood (QML) is an estimation technique, proposed by Heathcote, Brown, and Mewhort (2002), that provides robust and efficient estimates of distribution parameters, typically for response time data, in sample sizes as small as 40 observations. In view of the computational difficulty inherent in implementing QML, we provide open-source Fortran 90 code that calculates QML estimates for parameters of the ex-Gaussian distribution, as well as standard maximum likelihood estimates. We show that parameter estimates from QML are asymptotically unbiased and normally distributed. Our software provides asymptotically correct standard error and parameter intercorrelation estimates, as well as producing the outputs required for constructing quantile–quantile plots. The code is parallelizable and can easily be modified to estimate parameters from other distributions. Compiled binaries, as well as the source code, example analysis files, and a detailed manual, are available for free on the Internet.

Fitting parametric distribution models to empirical response time (RT) data has an importance that is increasingly recognized in quantitative psychology (e.g., Andrews & Heathcote, 2001; Balota & Spieler, 1999; Heathcote, Popiel, & Mewhort, 1991; Luce, 1986; Mewhort, Braun, & Heathcote, 1992; Ratcliff, 1978; Ratcliff & Murdock, 1976; D. Smith & Mewhort, 1998; P. L. Smith, 1995; Spieler, Balota, & Faust, 1996; Van Zandt, Colonius, & Proctor, 2000; Wixted & Roher, 1993). Recently, interest in psychology has focused not only on which parametric models should be used, but also on precisely how their parameters should be estimated from data (e.g., Cousineau & Larochelle, 1997; Heathcote, 1996; Van Zandt, 2000).

Estimating distributional parameters from empirical data is a difficult task, given the conditions that usually prevail in psychological measurement: high noise and small sample sizes. Various estimation methods have been proposed, most of which fall into two classes: either least-squares estimation based on sample statistics (e.g., quantiles, vincentiles, or cumulative densities) or maximum likelihood estimation. Van Zandt (2000) evaluated and compared most current estimation techniques and concluded that the standard maximum likelihood estimator (continuous maximum likelihood, CML) performed best. In addition, CML performs adequately on sample sizes as small as 100 observations, which is important for psychological applications (Ratcliff, 1979).

Recently, Heathcote, Brown, and Mewhort (2002) proposed a new variant of maximum likelihood estimation: quantile maximum likelihood (QML). Heathcote et al. (2002) evaluated the performance of QML estimation for data generated from ex-Gaussian distributions, in realistically small sample sizes. QML proved to be far superior even to CML estimation; QML parameter estimates were typically a little less biased and very much less variable than CML parameter estimates. Because of its improved efficiency, QML produced good parameter estimates in sample sizes as small as 40. This represents a very useful advance in methodology, given that RT sample sizes are often limited to well below 100 points because of such factors as the number of available experimental stimuli. QML is also relatively robust to the presence of outliers, whereas CML estimates are sensitive and, in some cases, are completely dominated by outliers. The robustness of QML makes it particularly attractive for use in psychological research. Note that these advantages of QML over CML cannot be expected to hold for all distributions and all sample sizes (e.g., the standard theorems tell us that these differences must either reverse or disappear given infinitely large sample sizes); however, preliminary simulation work suggests that QML's advantages extend to some other distributions commonly used to model RT data.

The QML estimation technique of Heathcote et al. (2002) suffers from two computational disadvantages: It is more complicated to implement and, in most cases, has

Correspondence concerning this article should be addressed to S. Brown, Department of Cognitive Sciences, University of California, Irvine, CA 92697-5100 (e-mail: scottb@uci.edu).

a higher computational cost than does CML. To reduce these problems and to take the burdens of calculating gradients and Hessians for the likelihood objective function from the end user, we present an open-source code Fortran 90 program for fitting the ex-Gaussian distribution by QML and CML methods. Although our program uses QML solely for the estimation of the ex-Gaussian distribution, there is no reason to limit this new technique to the ex-Gaussian. As such, we have developed the source code so that end users can easily alter it to implement QML estimation for other distributions.

Our program (QMLE, Quantile Maximum Likelihood Estimator) calculates maximum likelihood estimates of the parameters of the ex-Gaussian distribution from sets of observed data. Parameter estimates from the QML method are asymptotically (in the sense of increasing sample size) unbiased and normally distributed, like those of standard maximum likelihood. The QMLE program provides estimates of the standard errors and intercorrelations of the parameter estimates, based on the observed Fisher information matrix. These estimates are asymptotically accurate, as is shown in the Appendix.

In finite samples, we checked the agreement of the variance–covariance matrices provided by QMLE against values calculated by Monte Carlo simulation (with 500 repetitions for each sample size and parameter set). Using parameter settings identified by Heathcote et al. (2002) as "typical," we found that the mean standard deviations of the parameter estimates extracted from the variance–covariance matrices provided by QMLE agreed with the true values to within 4.5% for samples of size $N = 1{,}000$. Indeed, the asymptotic estimates were fairly accurate even in quite small samples: standard deviations of parameter estimates different from the true values by less than 7.5% for samples of size $N = 200$, and by less than 13.2% for samples as small as $N = 50$. Parameter intercorrelations extracted from the estimated variance–covariance matrices were likewise quite accurate in finite samples (agreement to within 5.3% for $N = 1{,}000$, 8.2% for $N = 200$, and 26% for $N = 50$). For further evaluation of the fits, QMLE returns observed and expected quantiles so that Q–Q plots (Cleveland, 1985) can be easily constructed.

QMLE was designed to run in a batch-processing mode, analyzing many different data sets contained in a single input file and requiring no user input after initiation. For greater flexibility, we have also included an interactive mode, in which the user can manually alter the parameters that affect the estimation procedure, such as the start point, convergence criteria, and parameters for the calculation of quantiles.

CML Versus QML Estimation for the Ex-Gaussian

For a full discussion of the details and merits of QML estimation, see Heathcote et al. (2002). We provide here only a brief comparison of this method with standard CML estimation. All maximum likelihood estimates are obtained by maximizing the likelihood of observed data over different theoretical parameter values. The likelihood function that is maximized depends on the statistical model assumed for the data. CML estimation of the ex-Gaussian assumes that the data are independent, identically distributed (i.i.d.) ex-Gaussian observations. Equation 1 gives the likelihood for the data under this model:

$$L(x_1, x_2, \ldots, x_N \mid \mu, \sigma, \tau) \propto \prod_{i=1}^{N} \mathrm{exg}(x_i, \mu, \sigma, \tau). \quad (1)$$

Here, we have assumed that the observations are $\{x_i: i = 1 \ldots N\}$ and that the ex-Gaussian probability density function (pdf) is expressed as Equation 2. Note that the likelihood is defined up to proportionality only, not up to equality, in accordance with Fisher's definition:

$$\mathrm{exg}(t, \mu, \sigma, \tau) = \frac{1}{\tau\sqrt{2\pi}} e^{-\frac{t-\mu}{\tau} + \frac{\sigma^2}{2\tau^2}} \int_{-\infty}^{\frac{t-\mu}{\sigma} - \frac{\sigma}{\tau}} e^{\frac{-y^2}{2}} dy. \quad (2)$$

QML estimation works by assuming a weaker model for the data, one that nests the model assumed by CML. This nesting ensures that QML estimation is always appropriate when CML estimation is appropriate and also allows QML estimation to work properly in the presence of conditions that would violate the assumptions of CML estimation. QML estimation begins with the specification of a fixed set of probabilities, $\{p_i: i = 0 \ldots m\}$, where $p_1 = 0$, $p_i < p_{i+1}$, and $p_m = 1$, such that $p(t < q_i) = p_i$, where the set $\{q_i: i = 0 \ldots m\}$ are quantiles. QML estimation assumes that the statistical model for the data is such that the frequency of observations in the interquantile ranges is equal to that which would be observed if the data were i.i.d. ex-Gaussian. This model leads to the likelihood function,[1] given as Equation 3:

$$L(x_1, x_2, \ldots, x_N \mid \mu, \sigma, \tau) \propto \prod_{i=1}^{m} \left(\int_{q_{i-1}}^{q_i} \mathrm{exg}(t, \mu, \sigma, \tau) dt \right)^{n_i}. \quad (3)$$

In practice, the quantiles are estimated from the observed data, and the $\{n_i: i = 1 \ldots m\}$ are integers, such that n_i represents the number of observations in the interval (q_{i-1}, q_i).

Equation 3 reveals why QML is significantly more costly than CML to implement for the ex-Gaussian pdf: The evaluation of the QML likelihood function requires many integrations of the ex-Gaussian pdf. This disadvantage will apply equally to any other pdf that is used, as long as an efficient analytic form for the cumulative density function (cdf) is not available. If the cdf does have a simple form, the integrals in Equation 3 can be replaced by differences of cdf values, and the numerical burden of QML estimation is reduced to approximately that of CML.

Equation 3 is not a fully specified likelihood function, if the $\{p_i\}$ are fixed and the $\{q_i\}$ are estimated from finite sample data.[2] This is because the $\{q_i\}$ are subject to sampling error and a fully specified likelihood function

must integrate Equation 3 over their joint sampling distribution. In fact, the joint sampling distribution of the $\{q_i\}$ follows the inverse of the incomplete beta function (see, e.g., Gilchrist, 2000, pp. 84–86) and a fully specified maximum likelihood approach would integrate across that distribution. The QML algorithm works in an approximate sense (and exactly in infinitely large samples) by replacing integration over this sampling distribution with a single representative value—the sample quantile. This approximation allows QML to work well, as compared with other methods, especially in small samples, as has been demonstrated by Heathcote et al. (2002). That QML is not a fully specified likelihood approach may trouble some readers, since the classic theorems of maximum likelihood estimation do not apply. However, the results established in the Appendix provide alternatives to the most basic theorems and allow QML users to be confident of its basic asymptotic properties.

The QMLE Program

The source code for the QMLE program can be obtained by following the links from http://www.newcastle.edu.au/school/behav-sci/ncl/ to the Software Repository page. In addition to the source code, we provide precompiled binaries for use with Windows and both parallel and serial 64-bit binaries for Solaris platforms. The Web site also contains a technical manual, which should be read before using QMLE. This manual provides detailed instructions for use, as well as covering the technical considerations we encountered in producing QMLE.

In writing QMLE, we have used only standard ANSI Fortran 90 code for portability; hence, the source code we provide should compile without modification on any platform with an ANSI-compliant compiler. The limitations of standard Fortran mean that QMLE has a quite primitive user interface. However, we feel that program portability and flexibility are more important than the interface. The only nonstandard program lines are some OpenMP compiler directives embedded in the slowest loops (OpenMP Architecture Review Board, 2000). Any noncompliant compiler will safely ignore these directives, but with an OpenMP compliant compiler, they will ensure that QMLE makes efficient use of symmetric multiprocessor hardware.

QMLE reads its operating parameters from a text file that contains all user-defined parameters and also allows for comments. In interactive fitting mode, the parameter file is reexamined, and changes are implemented between each search sequence, whereas in standard mode it is examined once only. The parameter file provides a record of the fitting conditions, including input and output file names, which can be enhanced by the insertion of comments that are ignored by QMLE. Sample parameter and data files are available for download with the program. Figure 1 (panel A, top left) gives an example of the format of the input file.

The data from the specified input file are analyzed in a cell-by-cell manner, with the data *cells* defined by a user-specified factor column, with the condition that data from the same cell form a contiguous block of lines in the data file. A sample of a data file appears in panel B (top right) of Figure 1: The left-hand column is the user-defined factor column that separates the cells to be analyzed; the right-hand column contains the data. The (compiled) program can then be called from a command shell and passed the name of any particular parameter control file the user wishes to employ. An example of the syntax used for this (suitable for both UNIX and Windows shells) appears in panel C (bottom) of Figure 1, followed by typical output generated by the program in the analysis of 14 cells of data (note that these output values, and more, are also written to files). The number of the cell being analyzed (i.e., its sequential position in the data file, and not necessarily the corresponding factor column value) is the first value on an output line, followed by the number of data points in that cell (N) and the number of quantile cut points calculated (M).

The analysis of each cell begins with the computation of those quantile estimates, for a user-defined set of quantile probabilities. Note that the calculation of these estimates is corrected for repeated observation values (runs) that often occur, due to limited-precision measurement. Our solution to this problem is to redistribute repeated values evenly across their range of possible true values. For example, suppose a value of 45 units was observed three times, with a measurement precision of ±0.5 units. Our algorithm redistributes these three observations as $\{44.75, 45, 45.25\}$ before estimating quantiles. We recognize that this solution is neither unique nor perfect but believe that it provides a workable resolution of the runs problem for experimental data. After correcting for runs, quantiles are estimated according to Hyndman and Fan's (1996) "Definition #5." Other users may prefer other quantile estimation algorithms, and so QMLE includes an option to operate on precalculated quantile estimates, rather than on raw data.

Once the quantile estimates are calculated, QMLE identifies a maximum for the likelihood function corresponding to either CML or QML, as directed by the user. The parameters for this maximum are the next three values on the output line (in the order μ, σ, and then τ). The start point for the search is generated using heuristics that we have found work well, even in small samples. However, we cannot guarantee the heuristics will work for all datasets, and so provision is made for user-defined start points. The search algorithm makes use of gradient values for the log-likelihood function during minimization. In Appendix A of the QMLE Technical Manual (available from the Web site), we provide generic expressions that calculate these gradients for both CML and QML, given values for the gradient of the underlying pdf. If the user wishes to estimate a pdf other than the ex-Gaussian, he or she needs only to replace the subroutines that provide gradient and pdf values for the ex-Gaussian distribution with routines appropriate for the new distribution (for full details, see the Technical

A: Parameter File Fragment

```
# Next line is input data file name.
sample.dat
# Next line is the output file stem
output
1       Measurement unit size.
1       Fitting Mode: 0=silent, 1=one output line/cell, 2=trace
# Parameters below here can be changed while fitting is running
1.e-9    Proportional objective function change tolerance
1.e-4    Proportional L(inf)-norm parameter tolerance
150      Maximum iterations allowed in search
2       Sample stat type (1=raw data, 2=quantile)
2       Data aggregation level
```

B: Data File Fragment

```
1     543
1     540
1     515
1     533
1     649
1     527
1     557
2     910
2     512
2     536
2     522
2     481
...
```

C: Example Session (Note: > is a command prompt, [↵] is an " Enter" keystroke)

```
> qmle [↵]
mypars [↵]
     1    39    31   445.63    56.45    72.17    26.20    19.38    28.44    3
     2    38    31   447.58    45.11   163.61    18.92    16.75    32.58    3
     3    40    31   478.15    10.52   122.50     8.20    11.99    21.93    1
     4    40    31   414.03     0.00   112.74     0.00     0.00     0.00   42
     5    40    31   434.15    10.03   124.73     6.46     7.66    21.36    3
     6    40    31   494.33    41.06   106.83    15.39    11.30    23.19    3
     7    40    31   422.59    26.60    90.50    12.80    11.38    19.72    1
     8    39    31   483.91    77.11    90.38    59.68    44.72    63.66    3
     9    38    31   478.49    51.69   120.19    20.76    17.21    28.56    3
    10    40    31   425.43     9.55   131.29     6.78     9.00    22.52    1
    11    40    31   469.77    41.61   188.27    18.34    15.74    36.04    2
    12    38    31   522.63    32.97    84.17    12.45     9.76    18.77    3
    13    39    31   478.31    75.91    88.43    37.83    27.57    40.11    3
    14    37    31   502.42    68.02   129.88    29.17    23.99    36.68    1
End of file reached in input data.  Stopping.

>
```

Figure 1. (A) Example of format of parameter input file. (B) Example of format of data input file. (C) Example of how to call the binary "qmle" from a shell and pass parameter file "mypars" for control, followed by the associated output.

Manual). Given such new routines, QMLE will use quadrature to integrate the pdf and the gradient functions and will use its own algorithms to calculate the gradient for the likelihood function.

Search convergence is controlled by two user-defined tolerance values: on the objective function value and estimated parameter values. After convergence, QMLE evaluates the inverse of the Hessian of the log-likelihood function at the convergence point to provide estimates of parameter standard errors (the next three values on the output) and intercorrelations (written to file only). Again, for distributions other than the ex-Gaussian, the user need only alter the subroutine that provides ex-Gaussian pdf Hessian values; QMLE will numerically integrate this function and calculate Hessians for the likelihood function from these integrals, as is discussed in Appendix A of the Technical Manual. QMLE also calculates expected quantile values for the estimated distribution, using a line search. These values are written to one output file (with default extension ".oe"), and parameter estimates, standard errors, parameter correlations, and an exit code are written to another file (with default extension ".par"). The exit code summarizes the properties of the fitting process (e.g., what type of convergence occurred) and parameter estimates (e.g., whether the Hessian was singular).

REFERENCES

ANDREWS, S., & HEATHCOTE, A. (2001). Distinguishing common and task-specific processes in word identification: A matter of some moment? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 514-544.

BALOTA, D. A., & SPIELER, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General*, **128**, 32-55.

CLEVELAND, W. S. (1985). *The elements of graphing data.* Monterey, CA: Wadsworth.

COUSINEAU, D., & LAROCHELLE, S. (1997). PASTIS: A program for curve and distribution analyses. *Behavioral Research Methods, Instruments, & Computers*, **29**, 542-548.

EDWARDS, A. W. F. (1972). *Likelihood.* London: Cambridge University Press.

FISHER, R. A. (1956). *Statistical methods and scientific inference.* Edinburgh: Oliver & Boyd.

GILCHRIST, W. G. (2000). *Statistical modeling with quantile functions.* London: Chapman & Hall/CRC.

HEATHCOTE, A. (1996). RTSYS: A DOS application for the analysis of reaction time data. *Behavior Research Methods, Instruments, & Computers*, **28**, 427-445.

HEATHCOTE, A., BROWN, S., & MEWHORT, D. J. K. (2002). Quantile

maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review*, **9**, 394-401.

HEATHCOTE, A., POPIEL, S. J., & MEWHORT, D. J. K. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin*, **109**, 340-347.

HYNDMAN, R. J., & FAN, Y. (1996). Sample quantiles in statistical packages. *American Statistician*, **50**, 361-365.

JEFFREYS, H. (1939). *Theory of probability.* Oxford: Oxford University Press, Clarendon Press.

LUCE, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* New York: Oxford University Press.

MEWHORT, D. J. K., BRAUN, J. G., & HEATHCOTE, A. (1992). Response time distributions and the Stroop task: A test of the Cohen, Dunbar, and McClelland (1990) model. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 872-882.

OPENMP ARCHITECTURE REVIEW BOARD (2000). *OpenMP Fortran application program interface, version 2.0.* Available at http://www.openmp.org/.

RATCLIFF, R. (1978). A theory of memory retrieval. *Psychological Review*, **85**, 59-108.

RATCLIFF, R. (1979). Group reaction time distributions and the analysis of distribution statistics. *Psychological Bulletin*, **86**, 446-461.

RATCLIFF, R., & MURDOCK, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review*, **83**, 190-214.

SMITH, D., & MEWHORT, D. J. K. (1998). The distribution of latencies constrains theories of decision time: A test of the random-walk model using numeric comparison. *Australian Journal of Psychology*, **50**, 149-156.

SMITH, P. L. (1995). Psychophysically principled models of visual simple reaction time. *Psychological Review*, **102**, 567-593.

SPIELER, D. H., BALOTA, D. A., & FAUST, M. E. (1996). Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 461-479.

VAN ZANDT, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review*, **7**, 424-465.

VAN ZANDT, T., COLONIUS, H., & PROCTOR, R. W. (2000). A comparison of two response time models applied to perceptual matching. *Psychonomic Bulletin & Review*, **7**, 208-256.

WIXTED, J. T., & ROHER, D. (1993). Proactive interference and the dynamics of free recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 1024-1039.

NOTES

1. Equation 3 suggests that QML bears strong similarities to χ^2 estimation. In the nomenclature usually used with χ^2, the $\{n_i\}$ are *observed frequencies* (O), and the

$$\int_{q_{i-1}}^{q_i} \exg(t,\mu,\sigma,\tau)dt$$

are *expected proportions* (E/N). QML optimization takes place on a logarithmic transformation of Equation 3, which gives the same objective function as that used for χ^2 estimation, except with the usual $(O-E)^2/E$ term replaced by $O(\log E - \log N)$. Optimizations over these two functions are not very dissimilar. In fact, Jeffreys (1939, pp. 146-148) has presented a Taylor approximation argument showing that the QML objective function is equal to a modified χ^2 function (with denominator O, rather than the usual E) up to a third-order error term and an irrelevant additive constant. The similarity to χ^2 estimation extends to robustness against measurement error: In particular, sample observations can be measured with error without altering the QML objective function, as long as the error is not large enough to move the quantile bounds.

2. We are thankful to an anonymous reviewer for pointing this out.

APPENDIX

Notation

Suppose we have data $\{x_i, i = 1, \ldots N\}$ that are i.i.d. samples drawn from a distribution with cumulative distribution function $F(t,\boldsymbol{\theta}^*)$, with parameter p-vector $\boldsymbol{\theta}^*$. Suppose further that we use the methods of QML to estimate the parameter vector from these data, resulting in an estimate $\hat{\boldsymbol{\theta}}$. QML estimation begins with a fixed (user-defined) set of quantile probabilities, $\{s_i, i = 1 \ldots M\}$, $\Sigma s_i = 1$. For each data set, these probabilities are used to estimate a set of quantile boundaries $\{q_i, i = 0, \ldots, M\}$, with $F(q_0,\boldsymbol{\theta}) = 0$ and $F(q_M,\boldsymbol{\theta}) = 1$ for all $\boldsymbol{\theta}$. Let $\{p_i, i = 0 \ldots (M-1)\}$ denote the number of observations from the sample that lie in each of the intervals (q_i, q_{i+1}), so that $\Sigma p_i = N$. For reasonable quantile estimation algorithms, including that used in the QMLE software (see Hyndman & Fan, 1996), the quantile estimates converge to the simple order statistic estimates with increasing sample size. That is, if o_a is the ath order statistic of the sample, then for sufficiently large N we have that $F(q_i,\boldsymbol{\theta}^*) = o_a/N$, and o_a/N converges to the sample value

$$\sum_{j=1}^{i-1} s_j$$

with increasing sample size.

Theorem 1 is very similar to the theorems put forth by Fisher (1956, p. 148; see also Edwards, 1972), except that it is trivially altered for QML, rather than for standard ML. Lemmas 1 and 2 are weaker versions of the analogous proofs for standard ML. Note, however, that both lemmas and the theorem apply only asymptotically. For standard ML, the equivalent results to Lemmas 1 and 2 apply in finite samples, whereas the final result (analogous to Theorem 1) applies asymptotically only.

LEMMA 1. *If L_Q denotes the QML likelihood function, then asymptotically,*

$$\frac{\partial}{\partial\theta_k}\left(\int L_Q\left(x_i, i = 1..N\middle|\boldsymbol{\theta}^*\right)\right) = 0$$

for $k = 1 \ldots p$, where the integral is over the sample space of the data.

PROOF OF LEMMA 1. By the definition of the QML likelihood function:

$$\frac{\partial}{\partial\theta_k}\left(\int L_Q\left(x_i, i = 1..n\middle|\boldsymbol{\theta}^*\right)\right) = \frac{\partial}{\partial\theta_k}\int\left(\prod_{i=0}^{M-1}\left(F\left(q_{i+1},\boldsymbol{\theta}^*\right) - F\left(q_i,\boldsymbol{\theta}^*\right)\right)^{p_i}\right). \tag{A1}$$

APPENDIX (Continued)

Given that L_Q is always greater than zero, it is sufficient to show that the desired equality holds for $\ln(L_Q)$; thus, we must show that

$$\frac{\partial}{\partial\theta_k}\left(\int \ln\left(L_Q\left(x_i, i=1..n\middle|\boldsymbol{\theta}^*\right)\right)\right) = \frac{\partial}{\partial\theta_k}\int\left(\sum_{i=0}^{M-1} p_i \ln\left(F\left(q_{i+1},\boldsymbol{\theta}^*\right) - F\left(q_i,\boldsymbol{\theta}^*\right)\right)\right) = 0. \tag{A2}$$

Asymptotically, the quantile estimates q_i have Dirac (point) distributions given by the order statistics, as was noted above; thus, integration over the sample space becomes evaluation at these points only. Using the linearity of the derivative operator and the chain rule applied to the logarithmic function,

$$\frac{\partial}{\partial\theta_k}\left(\int \ln\left(L_Q\left(x_i, i=1..n\middle|\boldsymbol{\theta}^*\right)\right)\right) = N\left(\sum_{i=0}^{M-1}\frac{p_i}{N}\frac{\dfrac{\partial F\left(q_{i+1},\boldsymbol{\theta}^*\right)}{\partial\theta_k} - \dfrac{\partial F\left(q_i,\boldsymbol{\theta}^*\right)}{\partial\theta_k}}{F\left(q_{i+1},\boldsymbol{\theta}^*\right) - F\left(q_i,\boldsymbol{\theta}^*\right)}\right). \tag{A3}$$

By the asymptotic properties of the quantile estimator discussed above, we have that $F(q_{i+1},\boldsymbol{\theta}^*) - F(q_i,\boldsymbol{\theta}^*) = (p_i/N)$ for sufficiently large N, giving

$$\frac{\partial}{\partial\theta_k}\left(\int \ln\left(L_Q\left(x_i, i=1..n\middle|\boldsymbol{\theta}^*\right)\right)\right) = N\left(\sum_{i=0}^{M-1}\left(\frac{\partial F\left(q_{i+1},\boldsymbol{\theta}^*\right)}{\partial\theta_k} - \frac{\partial F\left(q_i,\boldsymbol{\theta}^*\right)}{\partial\theta_k}\right)\right). \tag{A4}$$

However, the central terms in the sum of Equation A4 cancel, leaving only

$$\frac{\partial}{\partial\theta_k}\left(\int \ln\left(L_Q\left(x_i, i=1..n\middle|\boldsymbol{\theta}^*\right)\right)\right) = N\left(\frac{\partial F\left(q_M,\boldsymbol{\theta}^*\right)}{\partial\theta_k} - \frac{\partial F\left(q_0,\boldsymbol{\theta}^*\right)}{\partial\theta_k}\right). \tag{A5}$$

Grouping the derivatives,

$$\frac{\partial}{\partial\theta_k}\left(\int \ln\left(L_Q\left(x_i, i=1..n\middle|\boldsymbol{\theta}^*\right)\right)\right) = N\left(\frac{\partial}{\partial\theta_k}\left(F\left(q_M,\boldsymbol{\theta}^*\right) - F\left(q_0,\boldsymbol{\theta}^*\right)\right)\right). \tag{A6}$$

By definition, $F(q_M,\boldsymbol{\theta}) = 1$ and $F(q_0,\boldsymbol{\theta}) = 0$ for any $\boldsymbol{\theta}$. Thus, $F(q_M,\boldsymbol{\theta}) - F(q_0,\boldsymbol{\theta}) = 1$ for all $\boldsymbol{\theta}$, and hence, all partial derivatives of this value with respect to θ_k, $k = 1 \ldots p$ are zero, which completes the proof of Lemma 1.

Note that in the proof of Lemma 1, the assumption that the data are i.i.d. samples from F can be weakened for QML, but not for standard ML, as long as Equation A4 follows from A3, given the weaker model. This is equivalent to saying that the above theorem holds for models weaker than the i.i.d. model employed by standard ML, to the extent that the asymptotic distribution of the p_i under the weaker model is equal to that expected under the standard i.i.d. model.

LEMMA 2. *With the above notation, asymptotically, and for j,k = 1 . . p,*

$$\frac{\partial^2}{\partial\theta_k\partial\theta_j}\left(\int L_Q\left(x_i, i=1..N\middle|\boldsymbol{\theta}^*\right)\right) = 0.$$

PROOF OF LEMMA 2. The standard algebra of derivatives, for any function $f \neq 0$ gives

$$\frac{\partial^2 \ln f}{\partial\theta_j\partial\theta_k} = \frac{1}{f}\frac{\partial^2 f}{\partial\theta_j\partial\theta_k} - \frac{1}{f^2}\frac{\partial f}{\partial\theta_j}\frac{\partial f}{\partial\theta_k}.$$

Applying this to

$$f = \int L_Q$$

and using Lemma 1 to remove the first derivatives, asymptotically we have

$$\frac{\partial^2}{\partial\theta_k\partial\theta_j}\int L_Q\left(x_i, i=1..N\middle|\boldsymbol{\theta}^*\right) =$$

$$-\int L_Q\left(x_i, i=1..N\middle|\boldsymbol{\theta}^*\right) \times \sum_{i=0}^{M-1}\frac{p_i}{N}\frac{\left(\dfrac{\partial F\left(q_{i+1},\boldsymbol{\theta}^*\right)}{\partial\theta_k} - \dfrac{\partial F\left(q_i,\boldsymbol{\theta}^*\right)}{\partial\theta_k}\right)\left(\dfrac{\partial F\left(q_{i+1},\boldsymbol{\theta}^*\right)}{\partial\theta_j} - \dfrac{\partial F\left(q_i,\boldsymbol{\theta}^*\right)}{\partial\theta_j}\right)}{\left(F\left(q_{i+1},\boldsymbol{\theta}^*\right) - F\left(q_i,\boldsymbol{\theta}^*\right)\right)^2}. \tag{A7}$$

APPENDIX (Continued)

By the definition of L_Q, and recalling that asymptotically the integral reduces to evaluation at the order statistics versions of the quantiles, the right-hand side of Equation A7 becomes

$$-L_Q\left(x_i, i=1..N\middle|\boldsymbol{\theta}^*\right)\times \left(\sum_{i=0}^{M-1}\left(\frac{p_i}{N}\frac{\dfrac{\partial^2 F\left(q_{i+1},\boldsymbol{\theta}^*\right)}{\partial\theta_k\partial\theta_j}-\dfrac{\partial^2 F\left(q_i,\boldsymbol{\theta}^*\right)}{\partial\theta_k\partial\theta_j}}{F\left(q_{i+1},\boldsymbol{\theta}^*\right)-F\left(q_i,\boldsymbol{\theta}^*\right)}\right)-\sum_{i=0}^{M-1}\frac{p_i}{N}\frac{\left(\dfrac{\partial F\left(q_{i+1},\boldsymbol{\theta}^*\right)}{\partial\theta_k}-\dfrac{\partial F\left(q_i,\boldsymbol{\theta}^*\right)}{\partial\theta_k}\right)\left(\dfrac{\partial F\left(q_{i+1},\boldsymbol{\theta}^*\right)}{\partial\theta_j}-\dfrac{\partial F\left(q_i,\boldsymbol{\theta}^*\right)}{\partial\theta_j}\right)}{\left(F\left(q_{i+1},\boldsymbol{\theta}^*\right)-F\left(q_i,\boldsymbol{\theta}^*\right)\right)^2}\right). \tag{A8}$$

The first sum in Expression A8 is zero, by the same arguments as those used for the proof of Lemma 1. That is, asymptotically $F(q_{i+1},\boldsymbol{\theta}^*)-F(q_i,\boldsymbol{\theta}^*)=(p_i/N)$, and the summands reduce to the numerator elements. Canceling the central terms as before leaves only the second derivatives evaluated at q_M and q_0. Recall that $F(q_M,\boldsymbol{\theta})=1$ and $F(q_0,\boldsymbol{\theta})=0$ for any $\boldsymbol{\theta}$, by definition, and hence, the second derivatives are zero, giving

$$\frac{\partial^2}{\partial\theta_k\partial\theta_j}\int L_Q\left(x_i,i=1..N\middle|\theta^*\right)=-L_Q\left(x_i,i=1..N\middle|\theta^*\right) \times\sum_{i=0}^{M-1}\frac{p_i}{N}\frac{\left(\dfrac{\partial F\left(q_{i+1},\theta^*\right)}{\partial\theta_k}-\dfrac{\partial F\left(q_i,\theta^*\right)}{\partial\theta_k}\right)\left(\dfrac{\partial F\left(q_{i+1},\theta^*\right)}{\partial\theta_j}-\dfrac{\partial F\left(q_i,\theta^*\right)}{\partial\theta_j}\right)}{\left(F\left(q_{i+1},\theta^*\right)-F\left(q_i,\theta^*\right)\right)^2}. \tag{A9}$$

The multiplier in Equation A9, $L_Q(x_i, i=1..N|\boldsymbol{\theta}^*)$, approaches zero as sample size N increases. To see this, note from the definition of L_Q that it is can be expressed as the product of N numbers:

$$L_Q\left(x_i,i=1..N\mid\boldsymbol{\theta}^*\right)=\prod_{i=1}^{N}\left(F\left(q_{d(i)+1},\boldsymbol{\theta}^*\right)-F\left(q_{d(i)},\boldsymbol{\theta}^*\right)\right)$$

where the function $d:\{x_i,\ i=1..N\}\to\{0,1,\ldots,M-1\}$ is defined by $x_i\in(q_{d(i)+1},q_{d(i)})$. Asymptotically, each factor in the product is smaller than

$$s_{\max}=\max_{i=1..M}\{s_i\},$$

and so, asymptotically, the multiplier in Equation A9 is smaller than $(s_{\max})^N$. Since $0<s_{\max}<1$ by definition, this product must thus approach zero as N increases. Thus, the right-hand side of Equation A9 is asymptotically zero. This completes the proof of Lemma 2.

Given the results of Lemmas 1 and 2, the standard theorems of ML apply to QML, providing for the asymptotic normality of its parameter estimates, as well as showing that they are asymptotically unbiased and have an asymptotically exact variance–covariance matrix given by the inverse of the negative Hessian matrix of $\ln L_Q$. These arguments are reproduced below, as Theorem 1, but can also be found in Fisher (1956).

THEOREM 1. *In large samples, under repeated resampling of the data, the p-vector of parameter estimates,* $\hat{\boldsymbol{\theta}}$*, from the methods of QML are normally distributed and unbiased, with variance–covariance matrix given by* $-H_{\boldsymbol{\theta}^*}^{-1}$*, where* $H_{\boldsymbol{\theta}^*}$ *is the Hessian matrix for* $\ln L_Q$*, evaluated at* $\boldsymbol{\theta}=\boldsymbol{\theta}^*$.

PROOF OF THEOREM 1. We first show that the distribution of the gradient vector for $\ln L_Q$ in large samples is normal, with mean zero and variance–covariance matrix $-H$. By using only the definition of expectation and the chain rule for differentiation, the expected value of the jth element of the gradient vector for $\ln L_Q$ is given by

$$\int L_Q(\boldsymbol{\theta})\left.\frac{\partial\ln L_Q}{\partial\theta_j}\right|_{\boldsymbol{\theta}^*}=\int L_Q(\boldsymbol{\theta})\frac{1}{L_Q(\boldsymbol{\theta})}\left.\frac{\partial L_Q}{\partial\theta_j}\right|_{\boldsymbol{\theta}^*}=\int\left.\frac{\partial L_Q}{\partial\theta_j}\right|_{\boldsymbol{\theta}^*}.$$

But this value is asymptotically zero, by the result of Lemma 1. We now examine the Hessian of $\ln L_Q$. Two applications of the chain rule give

$$E\left(\frac{\partial^2\ln L_Q}{\partial\theta_j\partial\theta_k}\right)_{\theta^*}=\int\left(L_Q\frac{\partial^2\ln L_Q}{\partial\theta_j\partial\theta_k}\right)_{\theta^*}=\int\left(\frac{\partial^2 L_Q}{\partial\theta_j\partial\theta_k}\right)_{\theta^*}-\int\left(\frac{1}{L_Q}\frac{\partial L_Q}{\partial\theta_j}\frac{\partial L_Q}{\partial\theta_k}\right)_{\theta^*}. \tag{A10}$$

APPENDIX (Continued)

Lemma 2 implies that the first term on the right-hand side of Equation A10 is asymptotically zero:

$$E\left(\frac{\partial^2 \ln L_Q}{\partial\theta_j \partial\theta_k}\right)_{\boldsymbol{\theta}^*} = -\int\left(\frac{1}{L_Q}\frac{\partial L_Q}{\partial\theta_j}\frac{\partial L_Q}{\partial\theta_k}\right)_{\boldsymbol{\theta}^*}. \tag{A11}$$

Similarly, two further applications of the chain rule provide that

$$E\left(\frac{\partial \ln L_Q}{\partial\theta_j}\frac{\partial \ln L_Q}{\partial\theta_k}\right)_{\boldsymbol{\theta}^*} = \int\left(L_Q\frac{\partial \ln L_Q}{\partial\theta_j}\frac{\partial \ln L_Q}{\partial\theta_k}\right)_{\boldsymbol{\theta}^*} = \int\left(\frac{1}{L_Q}\frac{\partial L_Q}{\partial\theta_j}\frac{\partial L_Q}{\partial\theta_k}\right)_{\boldsymbol{\theta}^*}. \tag{A12}$$

From Equations A11 and A12, we thus have

$$E\left(\frac{\partial \ln L_Q}{\partial\theta_j}\frac{\partial \ln L_Q}{\partial\theta_k}\right)_{\boldsymbol{\theta}^*} = -E\left(\frac{\partial^2 \ln L_Q}{\partial\theta_j \partial\theta_k}\right)_{\boldsymbol{\theta}^*}. \tag{A13}$$

The left-hand side of Equation A13 is the expectation of the product of elements of the gradient vector for $\ln L_Q$. Since the expected value of these elements is uniformly zero (from Lemma 1 and above), this term is simply the covariance of the elements of the gradient vector. Furthermore, given that the function $\ln L_Q$ is the sum of many independent random variables, the central limit theorem implies that, asymptotically, it will be normally distributed. Thus, the gradient vector for $\ln L_Q$ is asymptotically normal, with mean zero and variance matrix given by Equation A13.

Now consider the sampling distribution of the QML parameter estimates, $\hat{\boldsymbol{\theta}}$, under repeated data sampling. We must make the extra assumption that the surface defined by $\ln L_Q$ for different parameters $\boldsymbol{\theta}$ is quadratic near the value of $\boldsymbol{\theta}^*$. The assumption of a quadratic log-likelihood surface will be asymptotically accurate. Given this assumption, the Hessian for $\ln L_Q$ is constant, say H, with (i,j)th element $H_{i,j}$. With a perfectly quadratic surface, a single step of a Newton–Raphson optimization will move between the sample parameter estimates, $\hat{\boldsymbol{\theta}}$, and the true value, $\boldsymbol{\theta}^*$, according to

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^* - H^{-1}\cdot\nabla \ln L_Q\big|_{\boldsymbol{\theta}^*}. \tag{A14}$$

Here, ∇ represents the gradient operator. From the results above, $\nabla \ln L_Q$ is (asymptotically) normally distributed, with mean zero and variance–covariance matrix $-H_{\boldsymbol{\theta}^*}$. From the usual properties of normal variables under linear transformation, it immediately follows from Equation A14 that $\hat{\boldsymbol{\theta}}$ is also asymptotically normal, with mean $\boldsymbol{\theta}^*$ and variance matrix $-H_{\boldsymbol{\theta}^*}^{-1}$. This concludes the proof of Theorem 1.

(Manuscript received June 29, 2001;
revision accepted for publication August 20, 2002.)