Simulation of Bayesian Posterior Distributions of Parameters of Constrained Models *

Ronald Schoenberg
Aptech Systems, Inc.
and
The University of Washington

August, 1997

*Comments or suggestions may be addressed to Ronald Schoenberg, Aptech Systems, Inc., 23804 SE Kent-Kangley Rd, Maple Valley, WA 98038, or by e-mail to rons@aptech.com.

Abstract

Constrained Maximum Likelihood (CML) is a new software module developed at Aptech Systems for the generation of maximum likelihood estimates of statistical models with general constraints on parameters. These constraints can be linear or nonlinear, equality or inequality. The software uses the Sequential Quadratic Programming method with various descent algorithms to iterate from a given starting point to the maximum likelihood estimates.

Standard asymptotic theory asserts that statistical inference regarding inequality constrained parameters does not require special techniques because for a large enough sample there will always be a confidence region at the selected level of confidence that avoids the constraint boundaries. Sufficiently large, however, can be quite large, in the millions of cases when the true parameter values are very close to these boundaries. In practice, our finite samples may not be large enough for confidence regions to avoid constraint boundaries, and this has implications for all parameters in models with inequality constraints, even for those that are not themselves constrained.

The usual method for statistical inference, comprising the calculation of the covariance matrix of the parameters and constructing t-statistics from the standard errors of the parameters, fails in the context of inequality constrained parameters because confidence regions will not generally be symmetric about the estimates. When the confidence region impinges on the constraint boundary, it becomes truncated, possibly in a way that affects the confidence limit. It is therefore necessary to compute confidence intervals rather than t-statistics.

Previous work (R.J. Schoenberg, "Constrained Maximum Likelihood", Computational Economics, 1997) shows that confidence intervals computed by inversion of the likelihood ratio statistics (i.e., profile likeihood confidence limits) fail when there are constrained nuisance parameters in the model. This paper describes the weighted likelihood bootstrap method of Newton and Raftery ("Approximate Bayesian inference with the weighted likelihood bootstrap", J.R. Statist. Soc. B, 56:3-48,1994). This method generates simulations of the Bayesian posterior of the parameters. Confidence limits produced from these simulations may be interpreted as Bayesian confidence limits.

KEY WORDS: Maximum Likelihood, Inequality Constraints, Bayesian Statistical Inference

1 Introduction

Most modern statistical models contain restricted parameters. These include stationarity constraints in time dependent processes, and constraints on the positiveness of variances such as the conditional variances in GARCH models.

Transformations of parameters and penalty methods have been customarily used to enforce constraints in statistical models. Convergence to a solution with these methods, however, has not always been reliable.

Han (1977) proposed the Sequential Quadratic Programming (SQP) method for the optimization of functions with general equality and inequality constraints. This method was applied to a statistical problem by Jamshidian, et al., (1993). Software implementations followed: Matlab's optimization toolbox, SAS's Proc NLP, and Aptech System's **CML**. **CML** is the first implementation of the SQP method explicitly for the maximum likelihood estimation of constrained statistical models.

2 CML

CML) is a set of procedures written in the GAUSS programming language (Schoenberg, 1995) for the estimation of the parameters of models via the maximum likelihood method with general constraints on the parameters

CML solves the general weighted maximum likelihood problem

$$L = \sum_{i=1}^{N} \log P(Y_i; \theta)^{w_i},$$

where N is the number of observations, w_i is a weight. $P(Y_i, \theta)$ is the probability of Y_i given θ, a vector of parameters, subject to the linear constraints,

$$A\theta = B,$$
$$C\theta \geq D,$$

the nonlinear constraints

$$G(\theta) = 0,$$
$$H(\theta) \geq 0,$$

and bounds

$$\theta_l \leq \theta \leq \theta_u.$$

$G(\theta)$ and $H(\theta)$ are functions provided by the user and must be differentiable at least once with respect to θ.

CML finds values for the parameters in θ such that L is maximized using the Sequential Quadratic Programming method. In this method the parameters are updated in a series of iterations beginning with a vector of starting values. Let θ_t be the current parameter values. Then the succeeding values are

$$\theta_{t+1} = \theta_t + \rho\delta,$$

where δ is a $K \times 1$ *direction* vector, and ρ a scalar *step length*.

Define

$$\Sigma(\theta) = \frac{\partial^2 L}{\partial\theta\partial\theta'},$$
$$\Psi(\theta) = \frac{\partial L}{\partial\theta},$$

and the Jacobians

$$\dot{G}(\theta) = \frac{\partial G(\theta)}{\partial\theta},$$
$$\dot{H}(\theta) = \frac{\partial H(\theta)}{\partial\theta}.$$

For the purposes of this exposition, and without loss of generality, we may assume that the linear constraints and bounds have been incorporated into G and H.

The direction, δ is the solution to the quadratic program

$$minimize\ \frac{1}{2}\delta'\Sigma(\theta_t)\delta + \Psi(\theta_t)\delta,$$

$$\begin{aligned} subject\ to\quad & \dot{G}(\theta_t)\delta + G(\theta_t) = 0, \\ & \dot{H}(\theta_t)\delta + H(\theta_t) \geq 0. \end{aligned}$$

This solution requires that Σ be positive definite (In practice, CML incorporates a slight modification in the quadratic programming solution which relaxes this requirement to non-negative definite).

In practice, linear constraints are specified separately from the G and H because their Jacobians are known and easy to compute. And the bounds are more easily handled separately from the linear inequality constraints.

The SQP method requires the calculation of a Hessian, Σ, and various gradients and Jacobians, Ψ, $\dot{G}(\theta)$, and $\dot{H}(\theta)$. **CML** computes these numerically if procedures to compute them are not supplied.

Descent Algorithms. The Hessian may be very expensive to compute at every iteration, and poor start values may produce an ill-conditioned Hessian. For these reasons alternative algorithms are provided in **CML** for updating the Hessian rather than computing it directly at each iteration. These algorithms, as well as step length methods, may be modified during the execution of **CML**.

The most reliable method given a good starting point is the Gauss-Newton method where the Hessian is directly computed. This method is time-consuming, however, especially if it must be computed numerically which is most often the case. **CML** also offers two quasi-Newton methods, the BFGS (Broyden, Fletcher, Goldfarb, and Shanno), and the DFP (Davidon, Fletcher, and Powell). These latter methods require far less computation and are generally more useful when the starting points is poor.

In practice a combination of the two methods is the most successful.

Line Search Methods. Define the merit function

$$m(\theta) = L + \max \mid \kappa \mid \sum_j \mid g_j(\theta) \mid - \max \mid \lambda \mid \sum_\ell \min(0, h_\ell(\theta)),$$

where g_j is the j-th row of G, h_ℓ is the ℓ-th row of H, κ is the vector of Lagrangean coefficients of the equality constraints, and λ the Lagrangean coefficients of the inequality constraints. The line search finds a value of ρ that minimizes or decreases

$$m(\theta_t + \rho\delta),$$

where ρ is a constant. Given θ and d, this is function of a single variable ρ. Line search methods attempt to find a value for ρ that decreases m. Several methods are available in **CML**, STEPBT, a polynomial fitting method, BRENT and BHHHSTEP, golden section methods, and HALF, a step-halving method.

3 Confidence Limits by Inversion

CML computes a covariance matrix of the parameters that is an approximate estimate when there are constrained parameters in the model (Gallant, 1987, Wolfgang and Hartwig, 1995). When the model includes inequality constraints, however, confidence limits computed from the usual t-statistics – dividing the parameter estimates by their standard errors – are incorrect because they do not account for boundaries placed on the distributions of the parameters by the inequality constraints.

For this reason, confidence limits must be calculated directly when the model contains constrained parameters. Two methods for such calculation are discussed here, by inversion of an appropriate statistic, and by simulation of the distribution of the parameter.

For further discussion of confidence limits by inversion of the likelihood ratio statistic see Cox (1974), Cook and Weisberg (1990), Meeker and Escobar (1995), Schoenberg (1997).

3.1 Problems with Confidence Limits using Inversion

It is well known that the distributions of the Wald and likelihood ratio statistics are modified when the true value of a constrained parameter being estimated is on a constraint boundary (Gourieroux, et al., 1982, Self and Liang, 1987, Wolak, 1991). In finite samples these effects occur in the *region* of the constraint boundary, specifically when the true value is within $\epsilon = \sqrt{(\sigma_e^2/N)\chi^2_{(1-\alpha,k)}}$ of the constraint boundary. This has consequences for the calculation of the confidence limits described in the previous sections.

We are concerned here with the unidimensional problem; the determining the confidence limits of one parameter in the model, leaving all other parameters as "nuisance" parameters. This problem can be divided into three cases,

(1) parameter constrained, no nuisance parameters constrained,

(2) parameter unconstrained, one or more nuisance parameters constrained,

(3) parameter constrained, one or more nuisance parameters constrained.

For case 1, when the true value is on the boundary, the statistics are distributed as a simple mixture of two chi-squares. Monte Carlo evidence presented below will show that this holds as well in finite samples for true values within ϵ of the constraint boundary.

For case 2, the statistics are distributed as weighted mixtures of chi-squares when the correlation of the constrained nuisance parameter with the unconstrained parameter of interest is greater than about .8. A correction for these effects is feasible. However, for finite samples, the effects on the statistics due to a true value of a constrained nuisance parameter being within ϵ of the boundary are greater and more complicated than the effects of actually being on the constraint boundary. There is no systematic strategy available for correcting for these effects.

For case 3, the references disagree. Gourieroux, et al. (1982) and Wolak (1991) state that the statistics are distributed as a mixture of chi-squares. However, Self and Liang (1987) show that when the distributions of the parameter of interest and the nuisance parameter are correlated, the distributions of the statistics are not chi-square mixtures.

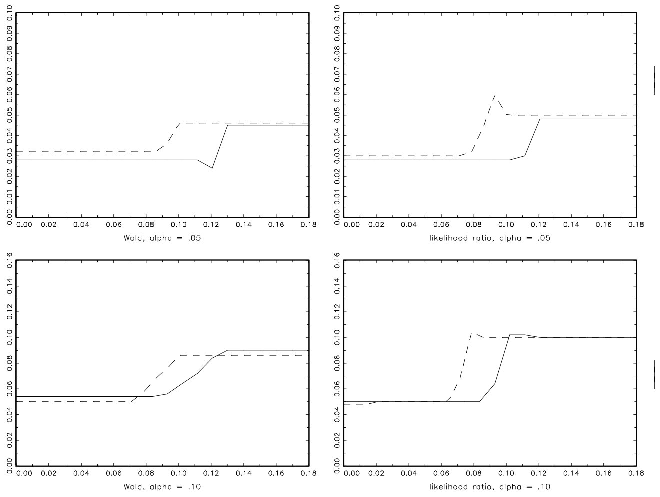

Figure 1: Size of constrained parameter of interest at different distances from the constraint boundary

3.1.1 Case 1: Confidence Limits of Constrained Parameter of Interest

A Monte Carlo analysis was conducted to explore the effects of a constraint boundary on the true size of the confidence limits computed by two methods, (1) inversion of the Wald statistic, and (2) inversion of likelihood ratio statistic.

The 95 percent likelihood ratio and Wald confidence limits for means constrained to be greater than zero were estimated for 40 models: a Normal with unit variance and 20 different true values for the means ranging from 0 to .18 for each of two sample sizes, 300 and 500.

The proportion of the confidence limits that failed to include the true value is plotted in Figure 1 against the true value. We observe that this proportion, or size, is about one half the correct size up to a threshold where it becomes the full correct size. The theoretical thresholds for confidence limits for a mean with Normal density with N = 300 and 500, at $\alpha = .05$, are .1131, .0876, and for $\alpha = .10$ are .0950, .0736, respectively. These threshold values are quite close to the Monte Carlo results in Figure 1.

3.1.2 Correction for Effects of True Value in Region of Constraint Boundary

The effects of having a true value near a constraint boundary may be corrected for by modifying the method of inversion of the chi-square statistic described in Section 3. For the likelihood ratio statistic, for example, a ϕ is found that satisfies

$$F_{lr}(\phi) = \begin{cases} \chi^2_{(1-2\alpha,1)}, & H(\tilde{\theta}) < \sqrt{(\sigma_e^2/N)\chi^2_{(1-\alpha,1)}} \\ \chi^2_{(1-\alpha,1)}, & H(\tilde{\theta}) \geq \sqrt{(\sigma_e^2/N)\chi^2_{(1-\alpha,1)}} \end{cases}$$

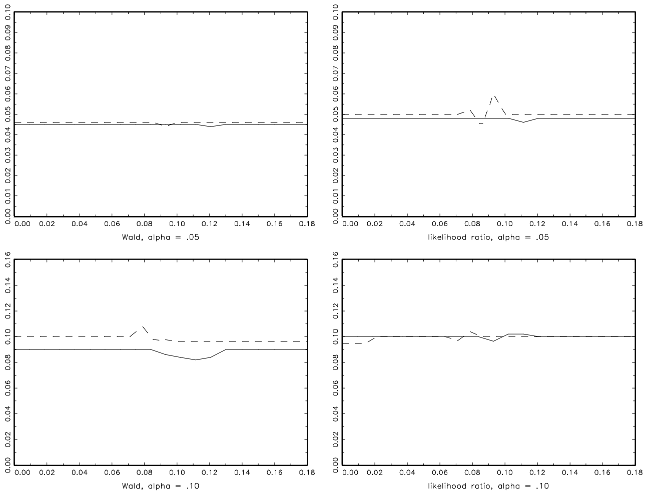

Figure 2: Size of constrained parameter of interest corrected

Upon request, CML applies this correction to the calculation of the confidence limits by inverting the chi square statistics. The Monte Carlo analysis reported in the previous section was repeated with this correction. Results displayed in Figure 2, where the success of the correction for the effects of the proximity of the constraint boundary may be observed.

3.2 Case 2: Confidence Limits of Unconstrained Parameter in Presence of Constrained Nuisance Parameters

When a constrained nuisance parameter is in the region of a constraint boundary, the confidence limits of a model parameter are affected even when the parameter is itself unconstrained. This fact is established when the true value of the nuisance parameter is on the boundary (Gourieroux, et al. (1982), Self and Liang (1987), and Wolak (1991)). When the parameter of interest and the nuisance parameter are uncorrelated, however, the effect vanishes.

A Monte Carlo study was conducted to determine possible effects of a constrained nuisance parameter in the region of a constraint boundary on the size of the confidence limits of an unconstrained parameter of interest. The true value of the constrained nuisance parameter varied from 0 to .18 in 37 intervals, and the correlation between the nuisance parameter and the constrained parameter of interest varied in 8 intervals from 0 to .999. 10,000 samples of size 500 were drawn under each of the conditions for a total of 2,960,000 samples. The results are presented in Figure 3.

The abscissa represents the true value of the constrained nuisance parameter, and the ordinate represents the observed size of the distribution of confidence intervals for the parameter of interest, that is, the proportion of intervals that fail to contain zero, the true value of the parameter of interest. The different curves show this relation at different correlations between the nuisance parameter and the parameter of interest.

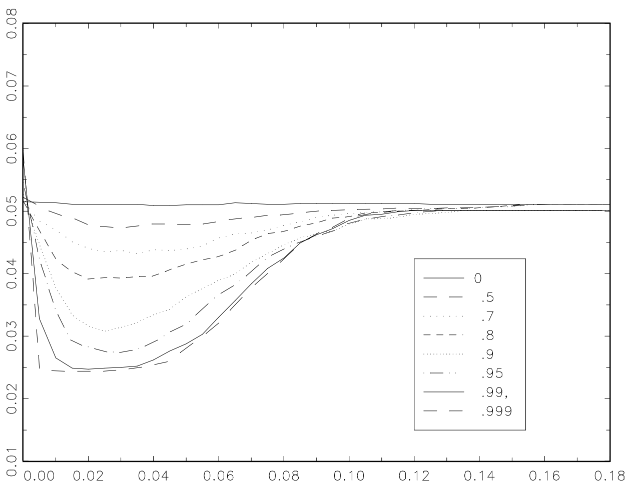

Figure 3: Size of confidence region of unconstrained parameter in the presence of a constrained nuisance parameter

We see from Figure 3 that any correction for size for a true value of the nuisance parameter on the boundary is not generalizable to true values in the region of the boundary, as it was for Case 1. In fact, effects on size in the region of the boundary are nonlinear and much larger than the effects on the boundary. This means that the corrections to the inversion of the chi-square statistics discussed in Gourieroux, et al. (1982) and Wolak (1991) are not applicable to cases where the true value of the correlated nuisance parameter falls within ϵ of the constraint boundary.

3.3 Case 3: Confidence Limits of Constrained Parameter in Presence of Constrained Nuisance Parameters

One should expect the behavior of the likelihood ratio statistic to be quite complex when both parameters are in the region of their boundaries, and this is confirmed in a Monte Carlo analysis.

The true values of the two means of a bivariate unit Normal distribution with correlation .9 were varied from 0 to .18 in 9 intervals. 500 samples of size 300 were drawn for each of these 225 sets of means. The 95 percent confidence intervals were computed by inversion of the likelihood ratio statistic, and the observed proportion of the intervals that failed to contain the true value are plotted in Figure 4.

Departures from nominal size are indicated by deviations from a flat plane set to .05. It is easily observed that relationship to nominal size is quite complex. There is no known method to compensate for this deviation. As with Case 2, however, departures from nominal size are trivial when the correlations between the parameter of interest and the nuisance parameter are less than about .7.

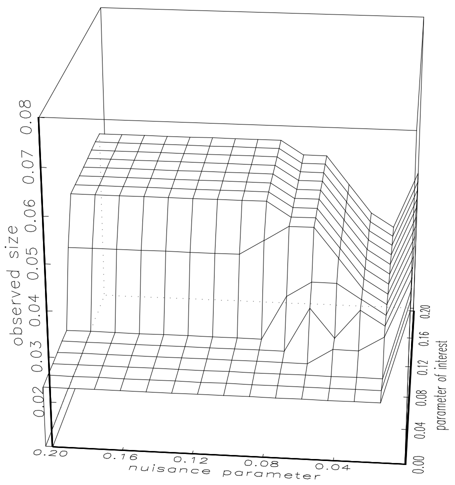

Figure 4: Size of confidence region of constrained parameter of interest in the presence of a constrained nuisance parameter

4 Confidence Limits by Simulation

As seen in the previous section, there is currently no general method for generating confidence intervals for models with constrained parameters with correct size using non-Bayesian methods. This conclusion is not modified for the simple bootstrap (Andrews, 1997). Newton and Raftery (1994), however, describe a weighted bootstrap that incorporates a prior distribution of the parameters that generates a simulated posterior distribution of the parameters for constrained models that appears to have correct size.

The weighted likelihood bootstrap is presented as an alternative to the Markov chain Monte Carlo methods (Geweke, 1995) for the model with constrained parameters

4.1 Weighted Maximum Likelihood Bayesian Simulation

CML procedure generates a simulation of the posterior distribution using the method of Newton and Raftery (1994).

In this method an SIR adjustment (Rubin, 1988) is made incorporating a prior to a weighted maximum likelihood using Dirichlet random variates for weights. The SIR weights are computed:

$$r(\hat{\theta}) = \pi(\hat{\theta})e^{L(\hat{\theta})}/\hat{g}(\hat{\theta}),$$

where $\pi(\hat{\theta})$ is the prior distribution of the parameters, and $\hat{g}(\hat{\theta})$ is a normal kernel density estimate of the joint density of the parameters using Terrell's (1990) method of maximum smoothing.

This method is quite easy to program. Moreover it is possible to write a general program that applies to any maximum likelihood estimation. Gibbs samplers and other Markov chain simulation methods require special programming for each type of model, and require substantially greater numbers of re-samplings. Experience suggests that reasonable accuracy is achieved with the weighted likelihood bootstrap on the order of the simple bootstrap. Thus for three places of accuracy, about 1000 re-samples are sufficient. Smaller numbers of re-samples will produce at least two places of accuracy, but the larger number of re-samples are generally required for kernel density plots.

4.1.1 Implementation in CML

First, a selected number, L, say, of weighted maximum likelihood estimations are produced where the weights are Dirichlet random variates. Next, we require an estimate of the covariance matrix of the parameters for the SIR weights. A natural candidate is the maximum likelihood moment matrix (Hartmann and Hartwig, 1995). Portions of this covariance matrix, however, are not available for those parameters whose maximum likelihood estimates are on constraint boundaries. Alternatively, an estimate could be constructed from the average estimate of the covariance matrix from the weighted likelihood re-samples. This is not ideal because, as with the unweighted maximum likelihood covariance matrix of parameters, some of portions of some of the re-sampled covariance matrices will not be available because the re-sampled estimates will be on constraint boundaries.

The implementation in CML computes an estimated covariance matrix directly from the re-sampled estimates. This is not ideal either since some of the estimates will be on constraint boundaries reducing variation. Fortunately, the SIR weights are quite robust in practice to choice of covariance matrix. In any event further research needs to be done on the estimate of the covariance matrix required for the SIR weights.

$r_i(\hat{\theta}), i = 1, 2 \cdots, L$ weights are calculated for the rows of the weighted bootstrap sample and normalized to sum to one. A Poisson random variate, $x_i =$ for each row is computed with mean $r_i(\hat{\theta})$. The final SIR adjusted sample is constructed by defining the frequency of the i-th row of the bootstrap sample as equal to x_i. Confidence limits and kernel density plots are generated from this sample.

4.2 Monte Carlo Study of Weighted Likelihood Bootstrap

In this Monte Carlo study of the size of the weighted likelihood bootstrap, random samples were generated from a unit normal distribution with true mean set to values ranging from zero to .12. Simple bounds were placed on the estimate of the mean restricting it to the [0,20] interval. Confidence limits were computed by the weighted likelihood bootstrap and by inversion of the Wald statistic. For the weighted likelihood bootstrap, a uniform prior was imposed for the interval [0,20].

For this study, the re-sample sizes were set to 100. This increases the magnitude of the stody by 100 times over that required for the study reported in section 3.1. The results for N = 500 and N = 800 are presented in Figure 5. The results for the inverse Wald follow the pattern established in section 3.1: the observed size is one half the theoretical size up to $\sqrt{(\sigma_e^2/N)} = .0876$ from the constraint boundary for N = 500, and .0693 for N = 800. At that point the correct size is restored.

The size of the confidence limits for the weighted likelihood bootstrap do not appear to be influenced by the constraint boundary which is what we are

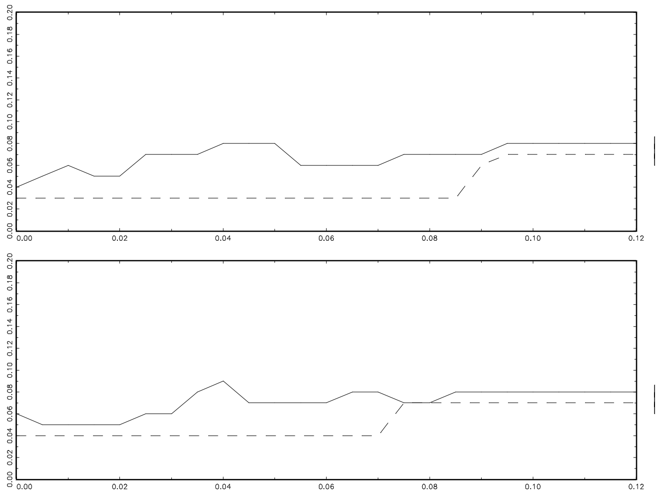

Figure 5: Size for Weighted Likelihood Bootstrap 95% Confidence Intervals, N = 500, 800

expecting. The curve is somewhat irregular, however, and it may be the case that re-sample sizes greater than 100 are required for some stability of size. A re-sample size of 500 is reasonable for a single analysis, but would put a Monte Carlo study out of practical range.

4.3 GARCH Model Example

The GARCH model for time series contains several highly constrained parameters. This example presents estimates and confidence limits for a GARCH(1,1) model applied to 20 years of monthly observations on the capitilization weighted returns of the Wilshire 5000 index.

Define the time series

$$\epsilon_t = y_t - \mu$$

where $t = 1, 2, ...T$, and y_t an observed time series with expected value μ.

Further define

$$\epsilon_t \equiv z_t \sigma_t$$

where $E(z_t) = 0$, $Var(z_t) = 1$, and

$$\sigma_t^2 = \kappa + \beta\sigma_{t-1}^2 + \alpha\epsilon_{t-1}^2;$$

The log-likelihood conditional on max(p,q) initial estimates of the conditional variances is, for $z_t \sim N(0,1)$

$$logL = -\frac{T-q}{2} log(2\pi) - \sum_{t+q}^{T} log(\sigma_t) - \frac{1}{2}\sum_{t+q}^{T}\frac{\epsilon^2}{\sigma_t^2}$$

The customary constraints applied to the parameters to enforce stationarity and a positive conditional variance are

$$\kappa > 0$$

Table 1: Parameter estimates and confidence limits by four methods.

		Standard Errors		inverse Wald		inverse LRS		Bayesian	
parameters	estimates	upper	lower	upper	lower	upper	lower	upper	lower
κ	1.1299	-0.4952	2.7550	0.0607	2.7550	0.2513	5.4624	0.1588	5.9648
β	0.8752	0.7802	0.9702	0.7802	0.9181	0.6549	0.9181	0.5682	0.9687
α	0.0719	0.0080	0.1359	0.0183	0.1148	0.0282	0.1148	0.0110	0.1421
μ	1.0947	0.6328	1.5566	0.6328	1.5566	0.6360	1.5571	0.5827	1.5841

Table 2: Estimated correlation matrix of the parameter estimates.

	κ	β	α	μ
κ	1.0000			
β	-.8221	1.0000		
α	-.0781	-.4688	1.0000	
μ	.0345	.0001	-.0506	1.000

$$\begin{aligned} \alpha+\beta &< 1 \\ \alpha &> 0 \\ \beta &> 0 \end{aligned}$$

To be able to construct a uniform prior, it is necessary to bound κ from above and μ from above and below. For the example,

$$\begin{aligned} \kappa &< 20 \\ \mu &> -20 \\ \mu &< 20 \end{aligned}$$

were added.

The prior, then, is constructed from the above constraints:

$$prior(\theta) = \begin{cases} V^{-1}, & \theta \in A \\ 0, & otherwise \end{cases}$$

where A is the hypervolume inscribed by the inequalities, and $V = 400$ for this example.

The parameter estimates are presented in Table 1 along with confidence limits computed by four different methods. It can be observed there that the confidence limits of the structural parameter, μ, is very consistent across the methods. μ is not itself constrained. However, it's confidence limits can be affected by estimates of the remaining parameters occuring in the region of their constraint boundaries. This, however, is provided that the distribution of μ is correlated more than about .7 with these parameters. Table 2 presents the estimate of the correlation matrix of the parameters calculated from the maximum likelihood estimate. We see there that the largest correlation is quite small, -.05, and thus we should expect that the constraints on the remaining parameters will have little effect on the structural parameter.

Kernel density plots were constructed for each of the parameters from the weighted likelihood bootstrap results (Figure 6). The posterior distribution of μ appears to be slightly skewed resulting in a slight difference between the mean and the mode. The mean of the distribution, 1.1167, is very close to the maximum likelihood estimate, 1.0947. The mode, however, appears to be about 1.2.

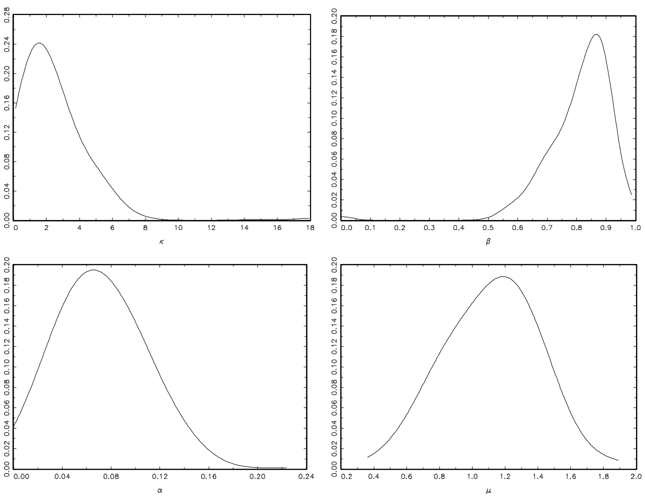

Figure 6: Kernel density plots of the parameters of a GARCH(1,1) model of the Wilshire 5000 capitalization weighted returns

5 Summary

Constraints are a common feature of modern statistical models. Statistical inference for models with constraints using the usual methods based on the Wald and likelihood ratio statistics turns out to be problematical. Because the constraints constitute prior information, though, Bayesian statistical inference would seem to be appropriate.

Geweke (1995) describes methods for generating simulations of posterior distributions of the parameters of constrained models. These methods, the Markov chain Monte Carlo simulators, are significantly more computationally intensive than even the simple bootstrap methods. Moreover, each statistical model requires special attention, obviating a general computer program.

Newton and Raftery (1994) propose an SIR adjusted weighted likelihood bootstrap for generating a simulated posterior distribution of the parameters. Monte Carlo evidence was presented here showing the relative success of the method.

6 References

Andrews, Donald W. K., 1997. "A simple counterexample to the bootstrap". *Cowles Foundation for Research in Economics*, Yale University.

Bates, Douglas M. and Watts, Donald G., 1988. *Nonlinear Regression Analysis and Its Applications.* New York: John Wiley & Sons.

Browne, Michael W. and Arminger, Gerhard, 1995. "Specification and estimation of mean- and covariance-structure models", in *Handbook of*

Statistical Modeling for the Social and Behavioral Sciences, Gerhard Arminger, Clifford C. Clogg, and Michael E. Sobel (eds.), New York: Plenum.

Cox, D.R., and Hinkley, D.V., 1974. *Theoretical Statistics.* London: Chapman and Hall.

Cook, R.D., and Weisberg, S., 1990. "Confidence Curves in Nonlinear Regression", *Journal of the American Statistical Association*, 85:544-551.

Meeker, W.Q., and L. A. Escobar, 1995. "Teaching about approximate confidence regions based on maximum likelihood estimation", *The American Statistician*, 49:48-53.

Dennis, Jr., J.E., and Schnabel, R.B., 1983. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Englewood Cliffs, NJ: Prentice-Hall.

Efron, Gradley, Robert J. Tibshirani, 1993. *An Introduction to the Bootstrap.* New York: Chapman & Hall.

Fletcher, R., 1987. *Practical Methods of Optimization.* New York: Wiley.

Gallant, A.R., 1987. *Nonlinear Statistical Models.* New York: Wiley.

Geweke, John, 1995. "Posterior Simulators in Econometrics", Working Paper 555, Research Department, Federal Reserve Bank of Minneapolis.

Gill, P. E. and Murray, W. 1972. "Quasi-Newton methods for unconstrained optimization." *J. Inst. Math. Appl.*, 9, 91-108.

Gourieroux, Christian, Holly, Alberto, and Monfort, Alain, 1982. "Likelihood ratio test, Wald Test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters", *Econometrica*, 50:63-80.

Han, S.P., 1977. "A globally convergent method for nonlinear programming." *Journal of Optimization Theory and Applications*, 22:297-309.

Hartmann, Wolfgang M. and Hartwig, Robert E., 1995. "Computing the Moore-Penrose inverse for the covariance matrix in constrained nonlinear estimation", SAS Institute, Inc., Cary, NC.

Hock, Willi and Schittkowski, Klaus, 1981. *Lecture Notes in Economics and Mathematical Systems.* New York: Springer-Verlag.

Jamshidian, Mortaza and Bentler, P.M., 1993. "A modified Newton method for constrained estimation in covariance structure analysis." *Computational Statistics & Data Analysis*, 15:133-146.

Newton, M.A. and Raftery, A.E., 1994. "Approximate Bayesian inference with the weighted likelihood bootstrap", *J.R. Statist. Soc. B*, 56:3-48.

O'Leary, Dianne P., and Rust, Bert W., 1986. "Confidence intervals for inequality-constrained least squares problems, with applications to ill-posed problems". *American Journal for Scientific and Statistical Computing*, 7(2):473-489.

Rubin, D.B., 1988. "Using the SIR algorithm to simulate posterior distributions", in *Bayesian Statistics 3*, J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith (eds.), pp. 395-402.

Rust, Bert W., and Burrus, Walter R., 1972. *Mathematical Programming and the Numerical Solution of Linear Equations.* New York: American Elsevier.

Schoenberg, Ronald, 1995. *CML Users Guide.* Maple Valley, WA, USA: Aptech Systems, Inc.

Self, Steven G. and Liang, Kung-Yee, 1987. "Asymptotic properties of maximum lieklihood estimators and likelihood ratio tests under nonstandard conditions", *Journal of the American Statistical Association*, 82:605-610.

Terrell, G.R., 1990. "The maximal smoothing principle in density estimation", *Journal of the American Statistical Association*, 85: 470-477.

Wolak, Frank, 1991. "The local nature of hypothesis tests involving inequality constraints in nonlinear models", *Econometrica*, 59:981-995.