CNA

CER Prediction Uncertainty

Lessons learned from a Monte Carlo experiment with an inherently nonlinear cost estimation relationship (CER)

Dr. Matthew S. Goldberg, CBO
Dr. Richard Sperling, CNA Corporation

April 30, 2007

Why we need prediction intervals

- Many cost estimation relationships (CERs) are nonlinear
- Often sample sizes are small, leading to uncertainty in predictions
- We need a method for quantifying the uncertainty when predicting from nonlinear CERs
  - Non-parametric Bootstrap
  - Delta method

Bootstrap vs. Delta method

- Bootstrap is a statistical re-sampling method (a variant of Monte Carlo) for estimating the bias and standard errors of statistical estimators with unknown sampling distributions
  - Re-samples with replacement from the empirical distribution of the prediction errors
  - Computationally intensive
  - Dr. Bradley Efron at UC Berkeley developed the method
  - Davison and Hinkley (1997) is a good book-length exposition
- Delta method is a *one-time* nonlinear calculation
  - Applies first-order Taylor series to the prediction, i.e., the CER
  - Then applies standard formula for the variance of a linear combination
  - Known since at least Tukey (1957) and Rao (1965)
- Validity of Bootstrap and Delta methods both based on asymptotic (large sample) statistical theory

A digression on asymptotics Freedman (2005, p. 168)

- "Statistical procedures are often defended on the basis of their 'asymptotic' properties—the way they behave when the sample is large…."
- "This is an oversimplification…."
- "Asymptotics are useful because they give clues to behavior for samples like the one you actually have."
- "Furthermore, asymptotics set a threshhold [*sic.*]. Procedures that do badly with large samples are unlikely to do well with small samples."

Design of our experiment

- Monte Carlo simulations of the Delta method and Bootstrap
- Assumptions
  - True CER of the "triad" form: $C = \alpha + \beta x^{\gamma}$
  - CER with multiplicative error: $C = (\alpha + \beta x^{\gamma}) \times \epsilon$
  - True parameter values: $\alpha = 1, \beta = 2, \gamma = 0.4$
  - $\epsilon \sim \mathrm{N}(1, 0.15)$
  - Assume cost driver (*x*) is uniformly distributed in [10,100]
  - Prediction for *x** = 80
- Repeat experiment for sample sizes (*n*) of 10, 25, 50, and 100

Algorithm for Monte Carlo experiment

For Monte Carlo replications 1 to 1,000:

1. Estimate CER by Iteratively Re-weighted Least Squares (IRLS)
2. Estimate mean and standard deviation of CER prediction across replications (the “truth”)
3. Estimate standard deviation of prediction using the Delta method at each replication
4. Estimate standard deviation of prediction using Bootstrap at each replication

Iteratively Re-weighted Least Squares

- Method for obtaining consistent (unbiased in large samples) parameter estimates and their standard errors for multiplicative-error models
- IRLS is a quasi-maximum likelihood estimator
- Goldberg and Touw (2003, pp. 63-5) show that for the multiplicative error CER that we deal with here, IRLS is equivalent to weighted nonlinear least squares, where the weights are updated each iteration, until convergence

Computational issues: Delta method

- Delta method requires only a single computational pass through the data
- Implemented as a single command in several statistical packages, e.g., Stata’s -nlcom- and -predictnl- routines
- Can be programmed in Excel, etc. using gradient, i.e., derivatives of CER function
  - In Excel, gradient must be programmed anew for each CER functional form
  - Not too difficult, but much easier in statistical programs

Illustration of the Delta method computation for the general case

- General nonlinear differentiable function: $F(Y_1 \cdots Y_m)$
- Gradient vector:
  - $\mathbf{g} = (\partial F/\partial Y_1 \cdots \partial F/\partial Y_m)$
- Delta method:
  - $\mathrm{Var}[F(Y_1, \ldots, Y_m)] \approx \mathbf{g}\,\mathrm{Cov}(\mathbf{Y})\,\mathbf{g}^T$

Illustration of the Delta method computation for a specific case

- Functional form: $F(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = \widehat{C} = \hat{\alpha} + \hat{\beta} x^{\hat{\gamma}}$
- Gradient vector: $\mathbf{g}^T = \begin{pmatrix} \partial\widehat{C}/\partial\hat{\alpha} \\ \partial\widehat{C}/\partial\hat{\beta} \\ \partial\widehat{C}/\partial\hat{\gamma} \end{pmatrix} = \begin{pmatrix} 1 \\ x^{\hat{\gamma}} \\ \hat{\beta} x^{\hat{\gamma}} \ln(x) \end{pmatrix}$
- Delta method: $\mathrm{Var}(\widehat{C}) \approx \mathbf{g}\, \mathrm{Cov}(\hat{\alpha}, \hat{\beta}, \hat{\gamma})\, \mathbf{g}^T$

$$= \begin{pmatrix} 1 & X^{\hat{\gamma}} & \hat{\beta} X^{\hat{\gamma}} \ln(X) \end{pmatrix} \left[\hat{\sigma}^2 \times \mathbf{V}\right] \begin{pmatrix} 1 \\ X^{\hat{\gamma}} \\ \hat{\beta} X^{\hat{\gamma}} \ln(X) \end{pmatrix}$$

Computational issues: Bootstrap

- Need to choose number of Bootstrap repetitions (*B*)
- For each repetition, re-sample with replacement from the empirical distribution and re-estimate the CER using IRLS
- Save results and tabulate
- Easier if done in statistical software
  - Still requires programming IRLS estimator for nonlinear CER functional forms
  - Seemingly not too hard—it's weighted nonlinear least squares
- We surmise that it is much more complicated to implement in Excel
  - Have to program IRLS estimator
  - Then iterate

Choice of number of Bootstrap repetitions

- Optimal number of bootstrap repetitions (*B*) is still an active topic of research
- Andrews and Buchinsky (2000) develop a method for choosing *B*
- Poi (2004) implements the method in Stata

Results of experiment

1,000 Monte Carlo replications		Sample size (n)			
		10	25	50	100
"True" parameters (Monte Carlo)	Mean prediction	?	12.5206	12.5278	12.5356
	SE of prediction	?	0.5278	0.3690	0.2810
Delta method	Avg value, SE of prediction	?	0.4907	0.3433	0.2642
Non-parametric bootstrap	Avg value, SE of prediction	?	?	?	?

Discussion of results

- All computations were done in version 9.2 of Stata (with assistance from Dr. Brian Poi of Stata Corp.)
- The Delta method provides reasonable estimates of the standard error of the prediction for *n*=25, 50, and 100
- Standard errors decline as the sample size increases
- No results for *n*=10 and Bootstrap due to IRLS convergence problems and time constraints
- Each reported result took several hours to obtain

Ill conditioning problem

- Recall that our CER is of the "triad" form with a multiplicative error
- IRLS applied to this functional form is equivalent to weighted nonlinear least squares
- Davidson and MacKinnon (1993, pp. 181-6) show that this estimation problem is ill-conditioned or "poorly identified"
- Nevertheless, may still be able to obtain parameter estimates for particular data sets
- There are two options when dealing with this problem
  - "Get more data"
  - "[E]stimate a less demanding model"
- Davidson and MacKinnon (1993, p. 186) write:

> If it is not feasible to obtain more data, then one must accept the fact that the data one has contain a limited amount of information and must simplify the model accordingly … **Only if the number of observations is very large, and/or the range of [the data] is very great, will it be feasible to obtain precise estimates of [the parameters] for this type of model** (emphasis added)

Summary

- In our experiment, the Delta method performs well for sample sizes of 25, 50, and 100 observations
  - Delta method may be inappropriate for some functional forms
  - Exploratory data analysis before statistical analysis
  - The Bootstrap is appropriate if you have sufficient observations and the functional form of the CER has extreme curvature and is not ill-conditioned
- Excel as an analytical tool
  - Not designed for iteration
  - Set-up time and run-time time are probably prohibitive for the Bootstrap
  - The Delta method is more straightforward to implement and provides reasonable estimates of standard errors for many nonlinear functional forms
- Bootstrap and Delta method are much easier when using a statistical program
  - The hardest task is to program the nonlinear IRLS estimator of the CER
  - After estimating the CER, type single commands for either Delta method or Bootstrap
  - Stata code available from me: sperlir@cna.org
- The IRLS estimator asks a lot of the data, too much in this case when the sample size is small

Final thoughts

- Sample sizes on the order of 10 or fewer observations are not uncommon when estimating CERs
- Given this fact and the result that the IRLS estimator of the triad form with a multiplicative error is ill-conditioned, why use the Bootstrap in this case?
- In general, you should *not* use a CER of the “triad” form with a multiplicative error unless you have a sufficient number of observations

References

- D.W.K. Andrews and M. Buchinsky. A three-step method for choosing the number of bootstrap repetitions. *Econometrica,* 68(1): 23-51, 2000.
- Russell Davidson and James G. MacKinnon. *Estimation and Inference in Econometrics*. Oxford University Press, New York, 1993.
- A.C. Davison and D.V. Hinkley. *Bootstrap Methods and their Application.* Cambridge University Press, New York, 1997.
- D.A. Freedman. *Statistical Models: Theory and Practice.* Cambridge University Press, New York, 2005.
- Matthew S. Goldberg and Anduin E. Touw. *Statistical Methods for Learning Curves and Cost Analysis.* Topics in Operations Research Series. INFORMS, Linthicum, MD, 2003.
- B.P. Poi. From the help desk: Some bootstrapping techniques. *Stata Journal*, 4(3): 312-28, 2004.
- C.R. Rao. *Linear Statistical Inference and Its Applications.* John Wiley, New York, 1965.
- John W. Tukey. The Propogation of Errors, Fluctuations and Tolerances: Basic Generalized Formulas. Technical Report 10, Princeton University, Statistics Research Group, December 1957.

That’s all folks

Questions?