Efficient Likelihood Ratio Confidence Intervals using Constrained Optimization

Gregor Reich Kenneth Judd

Econometric Society World Congress 2020

The Likelihood Ratio Test

- Setup
  - Model $\mathcal{M}$: Structural parameters $\theta \in \Theta$, states $x \in \mathcal{S}$, "outcomes" $y \in \mathcal{Y}$, policy/endogenous variables $\sigma \in \Sigma$
  - Model solution conditions $h(x; \sigma, \theta) = 0, \ \forall x \in \mathcal{S}$
  - Data set $\{\hat{x}_t, \hat{y}_t\}_{t=1}^T$
  - Log-likelihood function $L(\theta; \sigma) \equiv \log(P_{\mathcal{M}}(\{\hat{x}_t, \hat{y}_t\}_{t=1}^T; \sigma, \theta))$
- Estimation of θ (here: MPEC, but "nesting" NFXP):

$$\hat{\theta}, \hat{\sigma} = \arg \max_{\theta \in \Theta, \sigma \in \Sigma} L(\theta; \sigma)$$

$$\text{s.t. } h(x; \sigma, \theta) = 0, \ \forall x \in \mathcal{S}$$

- Likelihood ratio test
  - Hypothesis function: $\tau : \Theta \to \mathbb{R}$, $\tau \in \mathcal{C}^1$
  - Hypotheses: $H_0 : \tau(\theta) = 0$ against $H_1 : \tau(\theta) \neq 0$ (two-sided)
  - Test statistic: If H_0 is true, $2(L(\hat{\theta}; \hat{\sigma}) - L(\theta_0; \sigma_0)) \overset{a}{\sim} \chi_1^2$, where

$$\theta_0, \sigma_0 = \arg \max_{\theta \in \Theta, \sigma \in \Sigma} L(\theta; \sigma)$$

$$\text{s.t. } h(x; \sigma, \theta) = 0, \ \forall x \in \mathcal{S}$$

$$\tau(\theta) = 0$$

Test Inversion and Confidence Intervals

- Set of hypothesis values a which would *not* be rejected, given $L(\hat{\theta};\hat{\sigma})$

$$\mathcal{A}^\alpha \equiv \{a \in \mathbb{R} : \exists \theta, \sigma : h(x;\sigma,\theta) = 0 \text{ and } H_0 : \tau(\theta) = a \text{ not rejected at level } \alpha\}$$

  - Convex hull: $\mathcal{A}^\alpha \subseteq [\min(\mathcal{A}^\alpha), \max(\mathcal{A}^\alpha)] \equiv [\underline{a}, \bar{a}]$
  - $\mathcal{A} \neq \emptyset$ because $\tau(\hat{\theta}) \in \mathcal{A}^\alpha$; not a singleton if $L \in \mathcal{C}^0$ and $\alpha > 0$
- Computation of $\underline{a}$ ($\bar{a}$ analogously as max problem, or min $-\tau(\theta)$):

$$\begin{aligned} \hat{\underline{a}} = \min_{\theta \in \Theta, \sigma \in \Sigma} \; & \tau(\theta) \\ \text{s.t. } & h(x;\sigma,\theta) = 0, \ \forall x \in \mathcal{S} \\ & L(\theta;\sigma) \geq L(\hat{\theta};\hat{\sigma}) - 0.5\chi_1^2(1-\alpha) \end{aligned}$$

- $\mathcal{A}^\alpha$ forms a $(1-\alpha)\cdot 100\%$ confidence interval for $\tau(\theta)$
  - In repeated sampling experiments and estimations of θ, $\mathcal{A}^\alpha$ would contain the "true" value of θ in $(1-\alpha)\cdot 100\%$ of the times
  - "Duality of hypothesis testing and confidence intervals"
  - Dimension-wise confidence intervals of θ using $\tau : \theta \mapsto \theta_k$

The Bus Engine Replacement Model (Rust, 1987)

- Dynamic machine renewal problem
  - Payoff function

$$u(x, i; \theta) + \varepsilon(i) = \begin{cases} \theta_{RC} + \varepsilon(1) & i = 1 \\ \theta_1 \cdot x + \varepsilon(0) & i = 0 \end{cases}$$

  - Law of motion of the states:
    - $Pr(x'|x, i; \theta)$, with $Pr(x' < x|i = 0; \theta) = 0$ and $Pr(x' = 0|x, i = 1; \theta) > 0$
    - $\varepsilon \sim EV1$ i.i.d.
  - (Integrated) Bellman equation

$$\begin{aligned} EV(x, i) &\equiv \mathbb{E}[V(x', \varepsilon')|x, i] \\ &= \iint \max\{u(x', i'; \theta) + \varepsilon'(i') + \beta EV(x', i')\} Pr(x'|x, i; \theta) q(\varepsilon') d\varepsilon' dx' \\ &\equiv T[EV; \theta](x, i) \end{aligned}$$

- Estimate θ from data $\{x_t, i_t\}_{t,i}$ (here: MPEC, but "nesting" NFXP)

$$\hat{\theta}, \widehat{EV} = \arg \max_{\theta \in \Theta, EV} L(\theta; EV)$$

$$\text{s.t. } EV(x, i) = T[EV; \theta](x, i), \ \forall x \in \mathcal{S}, i \in \{0, 1\}$$

Confidence Intervals

- $(1-\alpha)\cdot 100\%$ Confidence intervals for $\tau = (\theta_{RC}, \theta_1, \theta_{RC}/\theta_1)$ (and $-\tau$)

$$\min_{\theta\in\Theta, EV} \tau_k$$
$$\text{s.t. } EV(x,i) = T[EV_\theta;\theta](x,i),\ \forall x \in \mathcal{S}, i \in \{0,1\}$$
$$L(\theta; EV) \geq L(\hat{\theta}; \widehat{EV}) - 0.5\chi^2_1(1-\alpha)$$

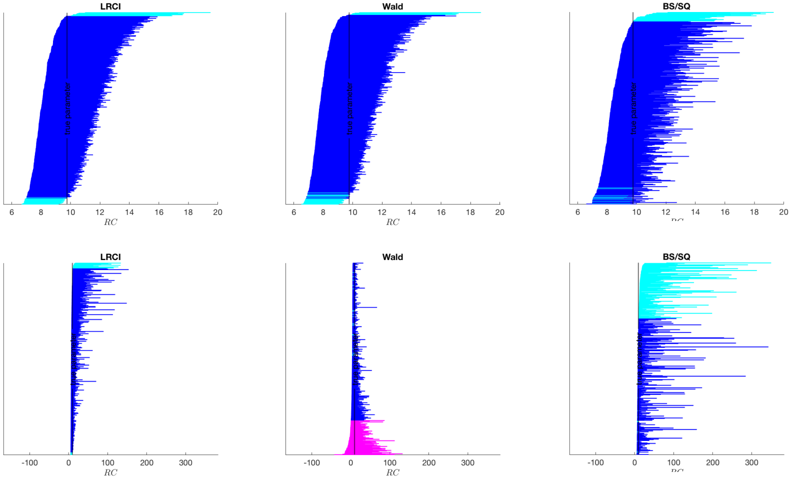
- Coverage analysis:
  - Simulate data sets under $\tilde{\theta}$
  - Estimate $\hat{\tilde{\theta}}$ and its confidence intervals
  - Check for inclusion of $\tilde{\theta}$
- Comparison:
  - Two different data set sizes (8,112 and 780)
  - Various types of confidence intervals
    - Likelihood ratio confidence intervals (LRCI)
    - Wald/SE (with delta method for mapped parameters)
    - Bootstrapping (sample quantiles)

Confidence Intervals: Coverage Analysis (1)

	LRCI					
	Sample size: 8,112			Sample size: 780		
	coverage	min	max	coverage	min	max
θ_{RC}	**0.961**	6.465	21.77	**0.958**	4.333	153.7
θ_1	**0.953**	0.558	7.888	**0.938**	7e-16	73.33
θ_{RC}/θ_1	**0.942**	2.348	12.07	**0.911**	1.305	4e07
	Wald/SE (with delta method)					
θ_{RC}	**0.952**	6.367	20.85	**0.955**	**-42.53**	132.8
θ_1	**0.928**	0.450	7.404	**0.935**	**-22.60**	61.00
θ_{RC}/θ_1	**0.962**	2.212	10.30	**0.791**	**-8e04**	8e04
	Bootstrap (sample quantiles)					
θ_{RC}	**0.928**	5.736	20.56	**0.675**	4.709	350.0
θ_1	**0.939**	0.273	7.723	0.813	1e-12	167.4
θ_{RC}/θ_1	**0.939**	2.231	11.11	0.880	1.181	5e12

	LRCI	Wald	Bootstrap
time (sec)	288	12	6,305

Confidence Intervals: Coverage Analysis (2)



Counter-Factuals: Demand Estimation in Rust (1987)

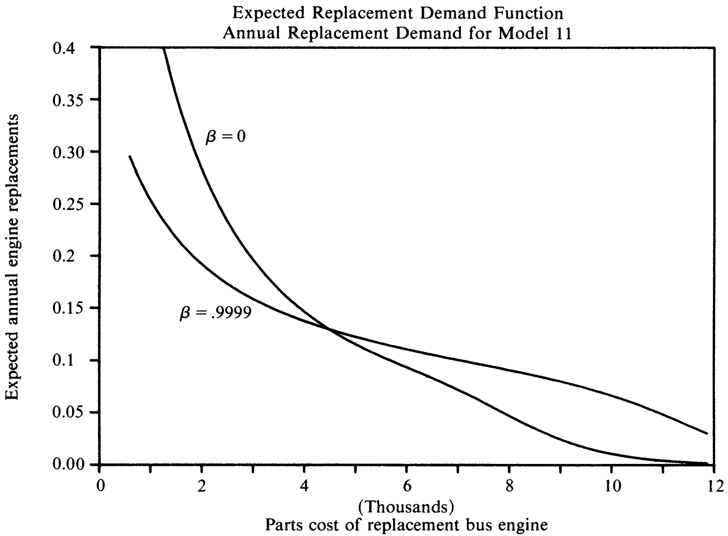
- Counter-factual: Use *estimated* model to carry out "policy experiments", e.g. by simulating/integrating the model variants to obtain and compare some derived quantity.
  - Assumption: Structural parameters are *policy-invariant*.
  - Goal: Analyze how estimation error propagates to derived quantities.
- **Counter-factual is a map of the parameters**, but its derivative is not always straightforward to compute (needed for delta method)
- Demand function estimation in Rust (1987)

$$d(\theta_{RC}) \equiv \int \pi_\theta(x, i = 1)dx$$

where the stationary distribution is defined as

$$\pi(x, i) = \iint Pr(i|x; EV_\theta)Pr(x|x', i'; \theta)\pi(x', i')dx'di',$$

Demand Curve in Rust (1987)

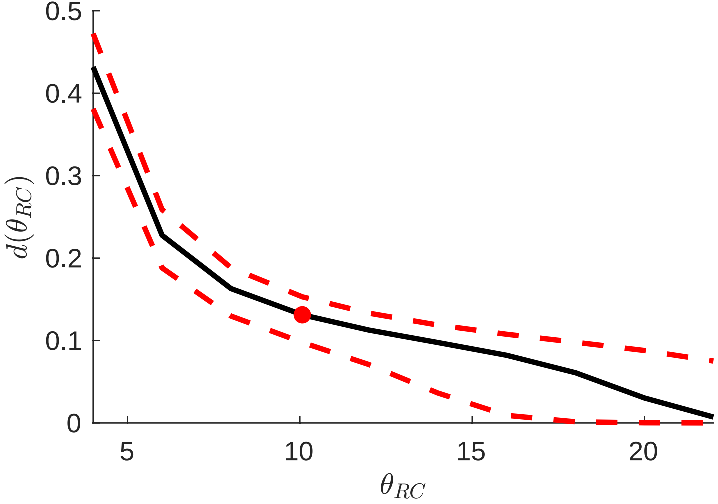


Confidence Intervals for Demand Curve (1)

- Confidence interval for $d(\theta_{RC})$ (θ_{RC} fix)

$$
\begin{aligned}
\hat{\underline{d}}(\theta_{RC}) = \arg \min_{\theta_1, \tilde{\theta}_{RC}, \pi, EV, \widetilde{EV}} & \int \pi(x, i=1) dx \\
\text{s.t. } \pi(x,i) = & \iint Pr(i|x; EV) Pr(x|x', i'; \theta_{RC}, \theta_1) \pi(dx', di'), \forall x, i \\
EV(x,i) = & \; T[EV; \theta_{RC}, \theta_1](x,i), \ \forall x, i \\
\widetilde{EV}(x,i) = & \; T[\widetilde{EV}; \tilde{\theta}_{RC}, \theta_1](x,i), \ \forall x, i \\
L(\tilde{\theta}_{RC}, \theta_1; \widetilde{EV}) \geq & \; L(\hat{\theta}; \widehat{EV}) - 0.5 \chi_1^2 (1-\alpha)
\end{aligned}
$$

Confidence Intervals for Demand Curve (2)



Conclusions

- We propose an efficient and easy-to-implement way to compute likelihood ratio confidence intervals (LRCI) for structural parameters—and mappings thereof—using constrained optimization
- We demonstrate that LRCI have very competitive coverage properties, in particular for mappings and smaller data sets; runtime performance is somewhere in between standard error based CIs and bootstrapping approaches
- We demonstrate the applicability to counter-factuals—a specific kind of mapping—which would otherwise be hard to assess for estimation error