Extrapolated empirical likelihood as a solution to the convex-hull-violation problem

Results from the working paper:
Kostyrka, A. V. (2025). Extrapolated empirical likelihood as a solution to the convex-hull-violation problem.

Andreï V. Kostyrka

UNIVERSITÉ DU LUXEMBOURG
Department of Economics and Management (DEM)

DEM internal seminar series
Faculty of Law, Economics, and Finance (FDEF)
University of Luxembourg
11th of November 2025

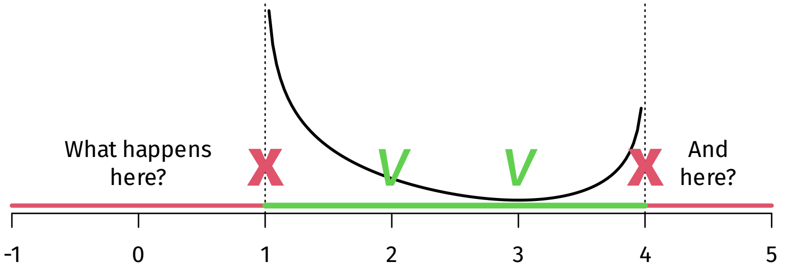
Presentation structure

1. Motivation and empirical applications
2. Empirical likelihood, its strong and weak points
3. Extrapolation of the EL ratio
4. Simulation results
5. Showcase of `smoothemplik`

Simplified explanation for my grandmother

We have a function that behaves well on some region but shoots off to infinity near a boundary, and is undefined beyond it.

How can we modify the exploding function so it remains usable where the original is undefined – perhaps by bending it smoothly?



Motivation and empirical applications

Contribution

I extend the literature and software ecosystem by:

1. Proposing a novel method of modifying the empirical likelihood (EL) so it returns sensible pseudo-values when the convex-hull condition is violated
2. Implementing two variants of this method and benchmarking their performance in numerical simulations
3. Releasing an open-source **R** package, **smoothemplik**, for fast non-parametric methods, including EL variants
   - First open-source implementation of smoothed EL for efficient estimation
   - Includes a range of modified EL variants from the literature

Motivating example

```
optim(theta0, smoothLik, method = "BFGS")
```

Error: initial value in ’vmmin’ is not finite

Motivation and research question

- Researchers often use EL to estimate model parameters θ from of moment conditions of the form $\mathbb{E}g(Z, \theta) = 0$
- This is done by maximising the EL ratio function (ELR)
- If zero lies outside the convex hull of $\{g(Z_i, \theta)\}_{i=1}^{n}$, then $\mathrm{ELR} = -\infty$
- Numerical optimisers cannot even start solving the model if the objective-function value is non-finite at the initial point

In plain terms: ‘*If all residuals have the same sign at some θ, EL optimisation fails*’.

How can we adjust the definition of the empirical likelihood ratio so that it is always finite?

Example from a micro-economic application

Angrist, J. D. Short-Run Demand for Palestinian Labor. *Journal of Labor Economics*, Vol. 14, No. 3 (Jul. 1996), pp. 425–453.

$$Y_i = Z'_{t(i)}\gamma + X'_i\beta + U_i,$$

where Z are endogenous macro aggregates (labour supply, wages, earnings) and X are included individual controls (dummies).

Estimate (γ, β) using Smoothed Empirical Likelihood (SEL; Kitamura, Tripathi & Ahn, 2004, *Econometrica*).

- Interactions of X forms 964 unique combinations (cells)
- Efficient estimation requires maximising the sum of local likelihoods over cells, but some cells are tiny (2–5 individuals)
- If model residuals U have the same sign in at least one cell, then SEL $:= -\infty \Rightarrow$ optimisation terminates prematurely

Literature review

Convex-hull failure: Grendár & Judge (2009, *EJS*); Bergsma, Croon & van der Ark (2012, *EJS*).

- Euclidean likelihood: Baggerly (1998, *Biometrika*); Owen (2001); Antoine, Bonnal, & Renault (2007, *JoE*)
- Penalised EL: Bartolucci (2007, *Stat. Prob. Lett.*); Lahiri & Mukhopadhyay (2012)
- Adjusted EL: Chen, Variyath, & Abraham (2008, *JCGS*); Liu & Chen (2010, *Ann. Stat.*)
- Balanced EL: Emerson & Owen (2009, *EJS*)
- Extended EL: Tsao (2013, *Canad. J. Stat.*); Tsao & Wu (2013, *Ann. Stat.*; 2014, *Biometrika*).

Empirical likelihood,
its strong and weak points

What is likelihood?

Likelihood: the probability of the observed data, viewed as a function of the fixed but unknown parameter θ.

Let $X_1, \ldots, X_n \overset{\text{IID}}{\sim} \mathcal{N}(\mu, 1)$. Then

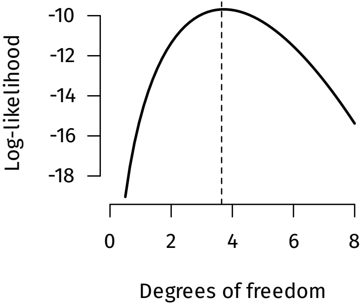
- $\mathcal{L}(\mu; X_1, \ldots, X_n) := \prod_{i=1}^n \phi(X_i - \mu)$
- $\ell(\mu; X_1, \ldots, X_n) = \log \mathcal{L}(\ldots) = -\frac{n}{2} \log 2\pi - \frac{1}{2} \sum_{i=1}^n (X_i - \mu)^2$
- The MLE is $\hat{\mu} = \bar{X}$

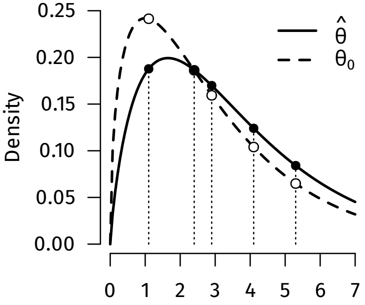
Likelihood illustration

Sample: $\{X_i\} = \{1.1, 2.4, 2.9, 4.1, 5.3\}$ drawn from a χ^2_3 distribution.

Likelihood: $\mathcal{L}(\theta; \{X_i\}_{i=1}^n) \colon \theta \mapsto f_{X;\theta}(X)$.

Log-likelihood: $\ell(\theta; \{X_i\}_{i=1}^n) = \sum_{i=1}^n \log f_{\chi^2_\theta}(X_i)$.





MLE: $\hat{\theta} = 3.65$. Likelihood: product of vertical line lengths.

Parametric and non-parametric likelihoods

- We may not know the true $f_{Y|X;\theta}$
- Models with a mis-specified $f_{Y|X;\theta}$ may perform acceptably (e. g. normal regression) or may fail (e. g. probit under heteroskedasticity, i. e. real-world conditions)

A non-parametric alternative aims to minimise assumptions:

- Assume $Y \mid X \sim F$, where F is
  - Continuous
  - Symmetric
  - Monotone
  - Convex

EL as a non-parametric likelihood

Empirical likelihood: Owen (1988, *Biometrika*; 2001).

- If we knew that $X \sim \mathcal{D}(\theta)$, we would maximise the log-likelihood
- Chamberlain (1987, *Econometrica*): any distribution can be approximated by a multinomial distribution with sufficiently many support points
- **Key idea:** replace the unknown CDF by the empirical CDF of the data
- The product of the mass-function values at $\{X_i\}$ is then the empirical likelihood $\mathrm{EL}(\mu)$

EL for the mean

For simplicity, begin with EL for the mean; the same logic extends to estimating equations under standard conditions.

Let $X_1, \ldots, X_n$ be the data and let the true mean be θ_0.

For any given θ, EL assigns probabilities $p_i > 0$, $\sum_i p_i = 1$ maximising $\sum \log p_i$ subject to $\sum_i p_i(X_i - \theta) = 0$.

Intuition: choose non-negative affine weights p_i so that the weighted average of $\{X_i\}_{i=1}^n$ equals θ, while making $\prod_i p_i$ as large as possible.

Profile EL ratio

Profile EL ratio (ELR) function: the weights $p_i > 0$ are profiled out.

$$\mathrm{ELR}(\theta) := \max_{p_i}\left\{\sum_{i=1}^{n} \log np_i \,\middle|\, \sum_{i=1}^{n} p_i = 1,\ \sum_{i=1}^{n} p_i X_i = \theta\right\}$$

Solution: $\hat{p}_i = n^{-1}(1+\hat{\lambda}'(X-\theta))^{-1}$, where $\hat{\lambda}$: $\sum_{i=1}^{n} \frac{X_i-\theta}{1+\hat{\lambda}'(X_i-\theta)} = 0$.

If $\{X_i\}$ are IID, with mean θ_0 and finite covariance of rank q,

$$-2\log \mathrm{ELR}(\theta_0) \xrightarrow[n\to\infty]{d} \chi^2_q.$$

under mild conditions (Wilks-type theorem).

Moment conditions in econometrics

Many applied models are built on conditional moment restrictions (CMR), typically in the form $\mathbb{E}(\rho(Z, \theta) \mid X) = 0$, where Z contains all included model variables and X comprises exogenous variables (instruments).

Examples: $\rho = Y - X'\theta$ for linear models; $\rho = Y - f(X, \theta)$ for non-linear models; $\rho = \mathbb{I}(Y - X'\theta < 0) - \tau$ for quantile regression.

CMRs are hard to work with $\Rightarrow$ in practice they are often converted into unconditional restrictions: $\mathbb{E}g(Z, \theta) = 0$, where $\dim g \geq \dim \theta$.

How can the model be solved without a GMM solver to feed g into?

EL for moment-based models

Estimating equations provide a way to estimate parameters in UCMR models. For the random variable $Z_{d\times 1}$, parameter $\theta_{p\times 1}$, and moment function $g_{q\times 1}$, suppose

$$\mathbb{E} g(Z, \theta_0) = 0$$

If $q = p$, then θ_0 is identified and the model is solved by setting $\frac{1}{n}\sum_{i=1}^n g(Z_i, \theta) = 0$.

If $q > p$, the specification can be tested and θ_0 may be identified. Estimate θ_0 by maximising $\mathrm{ELR}(\theta)$, i. e.

$$\max_{\theta} \max_{p_i} \left\{ \sum_{i=1}^{n} \log np_i \;\middle|\; \sum_{i=1}^{n} p_i = 1,\ \sum_{i=1}^{n} p_i g(Z_i, \theta) = 0 \right\}$$

EL for a linear model

$$Y = \alpha + \beta X + U = \tilde{X}'\theta + U, \quad \mathbb{E}(U \mid X) = 0$$

Estimating equation for θ:

$$\mathbb{E}\binom{1}{X}U = 0 \iff \mathbb{E}\binom{1}{X}(Y - X'\theta) = \mathbb{E}g(Z, \theta) = 0$$

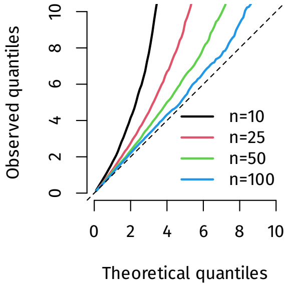
Here $U = Y - \tilde{X}\theta$ corresponds to $\rho(Z, \theta) = Y - \tilde{X}\theta$ – the moment function in the *conditional* restriction and $\binom{1}{X}$ are the instruments. Richer instrument sets: $(1, X, X^2, \exp X, |X|, \ldots)$ – yield over-identification.

For any θ such that $\mathrm{ConvHull}(\{g(Z_i, \theta)\}_{i=1}^{n}) \ni 0$, ELR can be computed using the procedure above.

EL calibration

Calibration: choosing critical values.

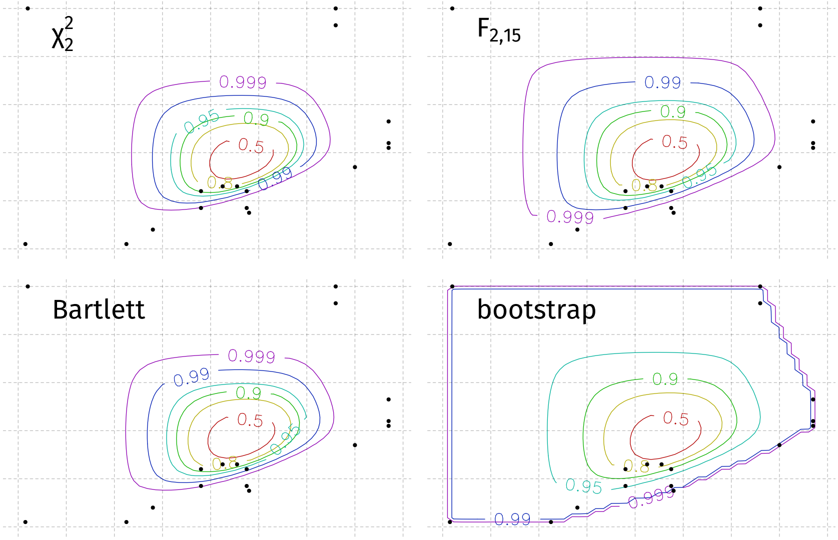
Unlike GMM (which relies on asymptotics), EL can be calibrated in finite samples!



Small *n*, non-zero skewness and excess kurtosis $\Rightarrow$ too-high observed values of $-2\log \mathrm{ELR}(\theta_0) \Rightarrow$ quantiles of $-2\widehat{\log \mathrm{ELR}}(\theta_0)$ lie *above* the χ^2 quantiles $\Rightarrow$ ELR test can over-reject, and EL can under-cover vs. the asymptotic baseline χ^2_p.

Therefore, the ELR statistic should be calibrated.

EL-based confidence regions



EL vs GMM

GMM drawbacks: requires an optimal weighting matrix, may produce implausible estimates, standard χ^2 tests rely on plug-in variance estimates.

EL advantages: internally studentised (no separate variance estimation), range-respecting, data-driven confidence-set shape, Bartlett-correctable.

- EL confidence sets are often more accurate at moderate *n*.

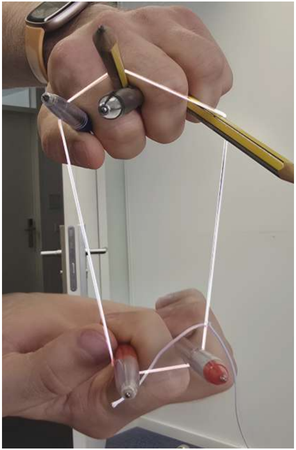
However: EL's convex-hull constraint can be a practical blocker – our focus today.

Convex hulls

For $\{X_i\}_{i=1}^n$, all possible combinations of $p_i X_i$ with $\sum_i p_i = 1,\ p_i \geq 0$ form the **convex hull** of the data.

Rubber-band analogy: stretching a rubber band (or applying a heat-shrink wrap) around the points creates the **boundary** of the convex hull.

In the EL context, all weighted means of the data with $p_i \geq 0, \sum_i p_i = 1$ lie in the convex hull of the data.



Convex-hull constraint

Feasibility requires 0 to lie within $\mathrm{ConvHull}\{g(Z_i, \theta)\}_{i=1}^n$. For the mean, θ must lie within $\mathrm{ConvHull}\{X_i\}_{i=1}^n$ because $g = X - \theta$.

- In one dimension (EL for the mean), at least one X_i must lie on the other side of θ – easy to check
- In higher dimensions this check becomes much harder: computing the convex hull has worst-case algorithmic complexity $O(n^{\lfloor q/2 \rfloor})$

If θ lies outside the convex hull, no $p_i \geq 0$ satisfy the moment constraint $\sum_i p_i g(Z_i, \theta) = 0$, and $\mathrm{ELR}(\theta) := -\infty$.

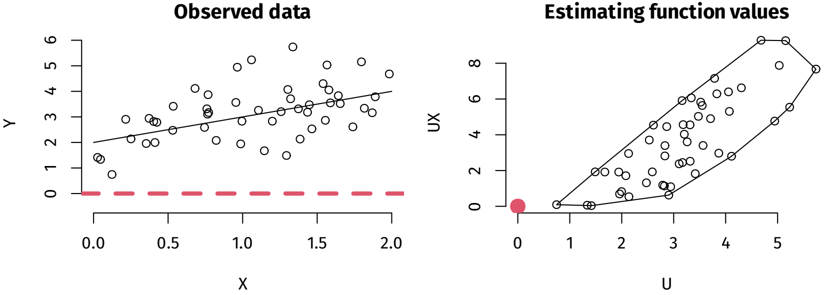
Why large data sets do not help

Convex-hull violations occur even when n is large (≈100,000).

- **Many dummies / categories:** with numerous indicators, moment conditions of the form $D\rho = 0$ arise; if $D = 1$ for a handful of observations, their residuals may share a sign
  - **Local smoothing:** local-likelihood methods with small bandwidths may fail when few observations carry weight
- **EL bootstrap:** resampling reduces distinct support (by $\approx 36.8\%$), increasing the chance that the sample mean lies outside the smaller bootstrap hull
- **Numerical optimisation:** in non-linear models, initialisation is often harder than optimisation; uninformed starts such as $(0, \ldots, 0)$ may force some coordinates of $g(Z, \theta)$ to have one sign

Convex-hull violation example 1

The data are positive; model estimation starts at (0, 0) ⇒ convex-hull violation!



Common in constrained optimisation: when maximising $\log \mathrm{ELR}(\theta^{(1)}, \theta^{(2)})$ w. r. t. $\theta^{(2)}$ (profiling $\theta^{(1)}$ as a nuisance), the admissible set for $\theta^{(1)}$ is very hard to guess *ex ante*.

Convex-hull violation example 2

A researcher bootstraps EL to improve small-sample inference. In many bootstrap samples the convex hull may exclude the full-sample average.

Film time!

Extrapolation of the EL ratio

Desirable properties of an EL modification

Any modification should preserve the shape of the original ELR where it behaves well and be the *minimal perturbation that fixed the failure*.

The modified EL should satisfy the following desirable properties.

- Be concave and smooth in θ (whenever the ELR is)
- Have Wald-type tails (be asymptotically χ^2)
- Admit standard calibrations (Bartlett, bootstrap etc.)
- Yield non-negative implied probabilities $\hat{p}_i$
- Be computationally as cheap as the original EL

Which properties are satisfied in our proposal?

Desirable properties of an EL modification

Any modification should preserve the shape of the original ELR where it behaves well and be the *minimal perturbation that fixed the failure*.

The modified EL should satisfy the following desirable properties.

- Be concave and smooth in θ (whenever the ELR is) ✓✓
- Have Wald-type tails (be asymptotically χ^2) ×,✓
- Admit standard calibrations (Bartlett, bootstrap etc.) ?,✓
- Yield non-negative implied probabilities $\hat{p}_i$ ×,×
- Be computationally as cheap as the original EL ✓,×

Which properties are satisfied in our proposal?

Wald parabola

EL can be approximated by a Wald quadratic.

Let $\bar{X}_n$ denote the sample mean and $\frac{1}{n}(\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)) = \frac{1}{n}\widehat{\mathrm{Var}}\, X$ its method-of-moments variance. Under the true null,

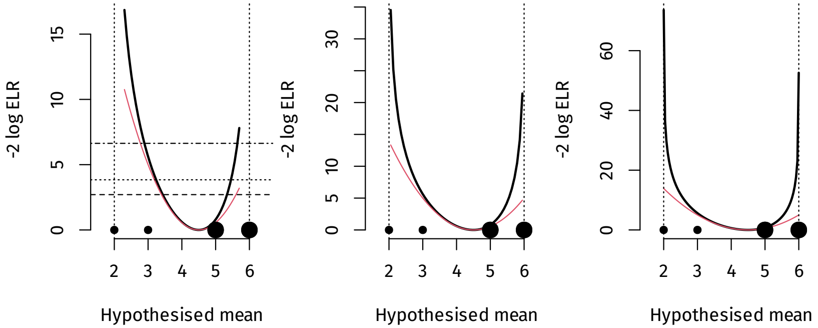
$$\hat{t}(\theta) := \frac{\bar{X}_n - \theta}{\sqrt{\frac{1}{n}\widehat{\mathrm{Var}}\, X}} \overset{n\to\infty}{\sim} \mathcal{N}(0, 1)$$

The square of this statistic forms a parabola in θ:

$$\hat{t}^2(\theta) = \hat{W}(\theta) := \frac{(\bar{X}_n - \theta)^2}{\frac{1}{n}\widehat{\mathrm{Var}}\, X} = n(\bar{X}_n - \theta)\widehat{\mathrm{Var}}^{-1} X(\bar{X}_n - \theta) \overset{n\to\infty}{\sim} \chi_1^2$$

For $\theta \approx \bar{X}_n$, $\log \mathrm{ELR}(\theta) \approx -0.5\hat{W}(\theta)$.

Explosive behaviour of EL



Thick black: −2 log ELR; red: Wald parabola. Horizontal lines: 90%, 95%, 99% confidence limits.

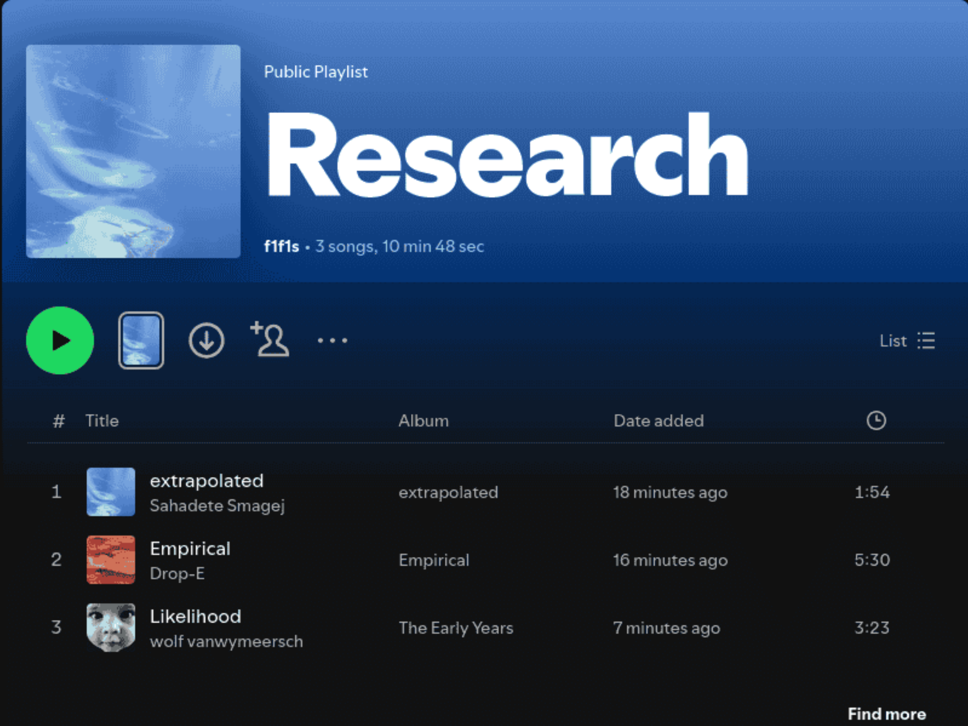
Extrapolating EL

Because $\log \mathrm{ELR}(\theta)$ diverges near the convex-hull boundary, we preserve it on a sufficiently large interior region and extrapolate beyond that region.

Extrapolated EL (ExEL): a modified $\log \mathrm{ELR}(\theta)$ such that

- Inside: $\mathrm{ExEL}(\theta) = \log \mathrm{ELR}(\theta)$
- Outside: ExEL replaces $\log \mathrm{ELR}(\theta)$ with a smooth, concave splice, ensuring finiteness

This separation allows the use of χ^2, F, Bartlett, and bootstrap calibrations.


Public Playlist
Research
f1f1s • 3 songs, 10 min 48 sec
List
Title
Album
Date added
1 extrapolated
Sahadete Smagej
extrapolated
18 minutes ago
1:54
2 Empirical
Drop-E
Empirical
16 minutes ago
5:30
3 Likelihood
wolf vanwymeersch
The Early Years
7 minutes ago
3:23
Find more

ExEL1 definition in one dimension

ExEL1: splice $\log \mathrm{ELR}(\theta)$ to its local quadratic (Taylor) expansion beyond a cut.

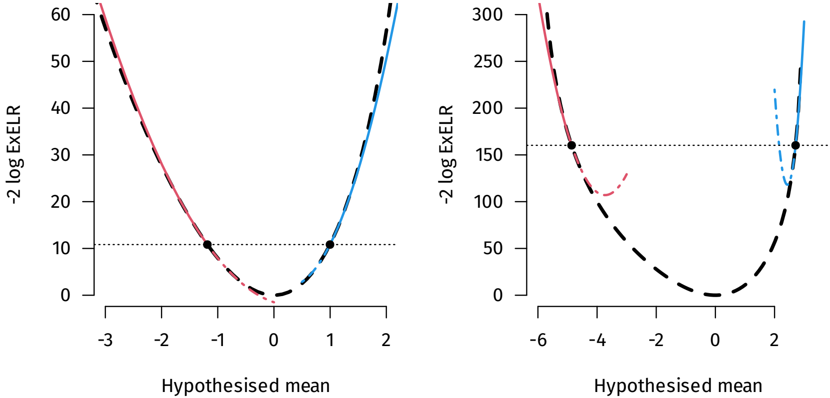
Right-tail procedure:

1. Choose $\theta_{\max} < X_{(n)}$
2. Compute $\log \mathrm{ELR}(\theta_{\max})$ and its first two derivatives
   - Closed-form expressions are available
3. Compute the coefficients of the parabola $a\theta^2 + b\theta + c$ matching value and first two derivatives at $\theta_{\max}$
   - With $f := \log \mathrm{ELR}$: $a = 0.5f''$, $b = f' - f''\theta_{\max}$, $c = f - f'\theta_{\max} + 0.5f''\theta_{\max}^2$
4. Use this quadratic for $\theta \geq \theta_{\max}$

Mirror for the left tail.

ExEL1 illustration

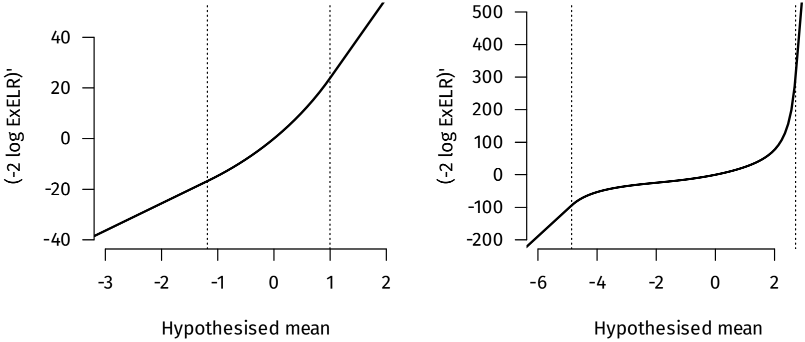
X: $-6, \ldots, 3$, counts (multiplicities): $1, \ldots, 10$.



Left: cut at $Q_{\chi^2_1}(0.999)$, right: cut at $Q_{\chi^2_1}(1 - 10^{-36})$.

ExEL1 smoothness

In EL for the mean, log ELR is concave and at least $\mathcal{C}^2$; values and slopes match at $\theta_{\max}$.



Left: cut at $Q_{\chi^2_1}(0.999)$, right: cut at $Q_{\chi^2_1}(1 - 10^{-36})$.

ExEL2 definition in one dimension

ExEL2: splice $\log \mathrm{ELR}(\theta)$ to the global Wald parabola via a smooth, concave exponential bridge $\alpha_0 + \alpha_1\theta + \alpha_2 \exp(\theta - \theta_{\max})$.

The tangency point θ^* on the Wald curve W is unknown and must be found.

1. Choose $\theta_{\max} < X_{(n)}$ (as for ExEL1)
2. Find $\theta^* := \theta_{\max} + t$ with $t > 0$ by solving

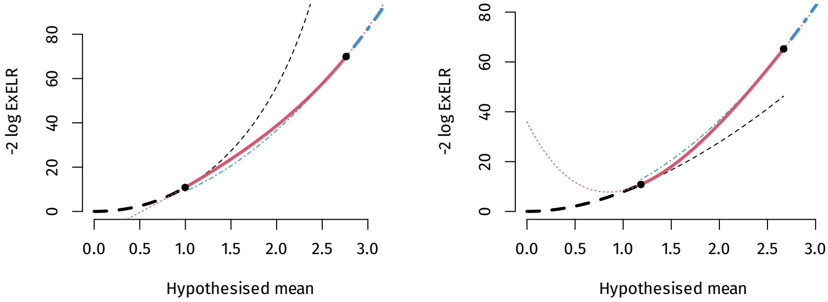
   $$G(t) := f'(\theta_{\max}) \cdot t + \frac{W'(\theta^*) - f'(\theta_{\max})}{e^t - 1}(e^t - 1 - t) - W(\theta^*) + f(\theta_{\max})$$

   and recover the bridge coefficients
3. Use the bridge for $\theta_{\max} < \theta < \theta^*$ and $W(\theta)$ for $\theta \ge \theta^*$

Mirror for the left tail.

ExEL2 illustration

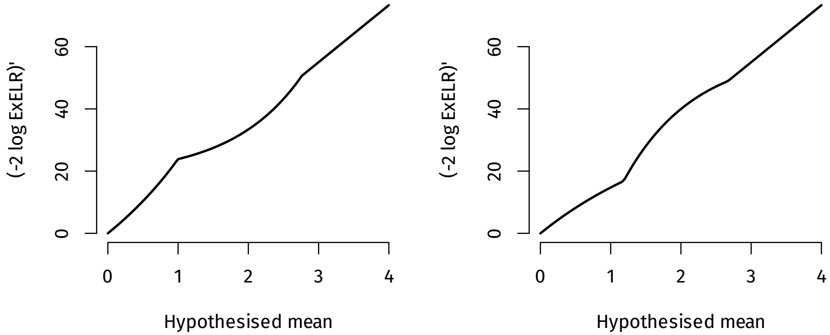
$X: -6, \ldots, 3$, multiplicities: $1, \ldots, 10$ (left);
$X: -3, \ldots, 6$, multiplicities: $10, \ldots, 1$ (right).



Cut at $Q_{\chi^2_1}(0.999)$.

ExEL2 smoothness

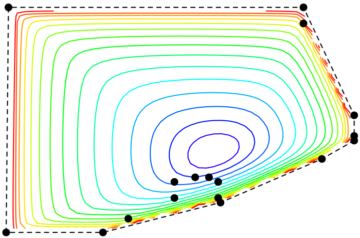
ExEL2 also has continuous derivatives: values and slopes match at θ_{max}. Left/right plots show the derivatives of the left/right functions from the previous slide.

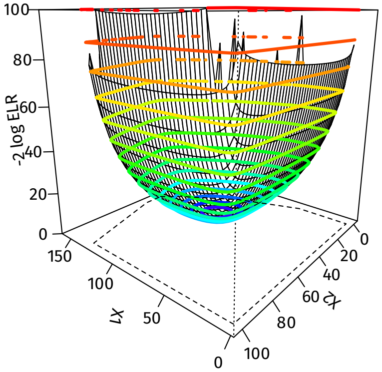


Choosing the cut-off

- Horizontal (1D): use a data-driven rule (e. g. $X_{(n-1)}$ or one median observation gap away from the boundary)
  - Or supply a custom θ_{cut} based on substantive considerations (e. g. extreme data quantiles)
- Vertical (in any dimension): use a numerical search (e. g. choose θ_{cut} such that $-2 \log \text{ELR}(\theta_{\text{cut}}) = Q_{\chi^2_{\dim \theta}}(1 - \alpha)$ with $\alpha = 0.001$)

EL ratio in multiple dimensions



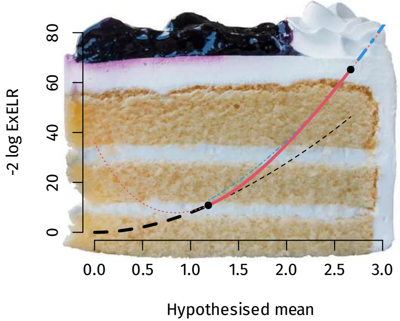


log ELR diverges as the putative θ approaches the convex-hull boundary.

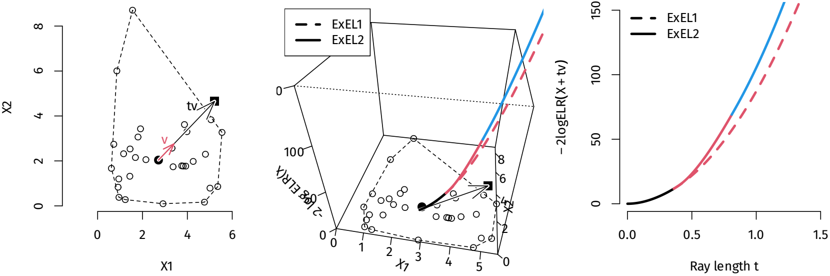
Radial reduction in multiple dimensions

View $-2 \log \text{ELR}(\theta)$ as a cake sliced radially from the centre towards θ. We then extrapolate along the 1D ray in the direction of θ.

$\Rightarrow$



Intuition in two dimensions



ExEL1 and ExEL2 in multiple dimensions

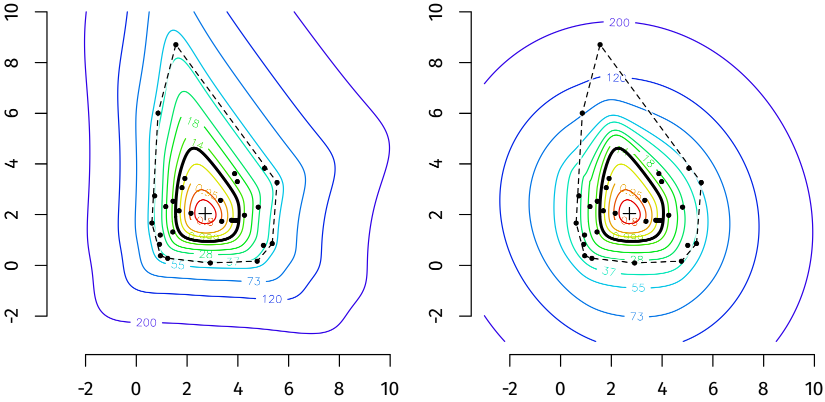
For any θ, define the direction from the sample average $v = \frac{\theta - \bar{X}_n}{\|\theta - \bar{X}_n\|}$ and the radius $t_\theta = \|\theta - \bar{X}_n\|$, to that $\bar{X}_n + tv = \theta$.

Work with the radial section $f_v(t) = \log \mathrm{ELR}(\bar{X}_n + tv)$.

ExEL1: find $t_{\max}$ with $f_v(t_{\max}) = Q_{\chi^2_p}(0.999)$ (or similar); compute f_v, f_v', f_v'' at $t_{\max}$ and fit the quadratic.

ExEL2: after choosing $t_{\max}$, find t^* so the bridge matches value and slope at $t_{\max}$ and is tangent to the Wald curve. The Wald section is $W_v(t) = -0.5n(v'(\widehat{\mathrm{Var}}\,\bar{X}_n)^{-1}v)t^2$.

Full 2D ExEL1 and ExEL2



Simulation results

Setting 1: ExEL helps in bootstrapping EL

The critical values of $-2 \log \mathrm{ELR}(\theta_0)$ in practice lie *above* those of χ^2_p $\Rightarrow$ under-coverage and over-rejection in finite samples.

To gauge the extent of inflation, once can bootstrap the statistic under $\mathcal{H}_0$: test that the bootstrap mean equals the observed sample average.

However, because bootstrap samples contain roughly 37% fewer distinct points, the convex hull of a bootstrap sample may exclude the sample average, *especially for skewed or heavy-tailed distributions*.

Redrawing 'until the convex-hull condition is satisfied' is inappropriate because it distorts sampling uncertainty.

ExEL is finite everywhere, so bootstrapping always yields a valid reference law.

Bootstrap percentile calibration

Resample the data B times, compute $\{-2 \log \mathrm{ELR}^{*(b)}(\bar{X}_n)\}_{b=1}^{B}$.

Reject if the observed $-2 \log \mathrm{ELR}(\theta_0)$ exceeds the bootstrap $(1-\alpha)$-quantile.

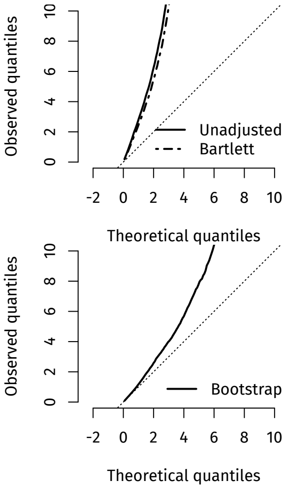
Unlike the bootstrap-*t*, which requires variance estimation (not obvious!), EL bootstrap can be as simple as

```
replicate(1000,
  EL(sample(X, length(X), replace = TRUE),
     mu = mean(X))$logelr)
```

Why ExEL helps calibration

The bootstrap reproduces sampling uncertainty, but with only two (or one) distinct points, uncertainty becomes excessive. This arises, as noted, in methods with many small cells containing few observations.

Illustration: for $n = 10$, the finite-sample quantiles of $-2 \log \text{ELR}$ are almost twice those of χ^2_1. The bootstrap provides a partial remedy; Bartlett is of limited help.



Simulation 1 design and workflow

1. Draw a random sample $\{X_i\}_{i=1}^n$, for $n = 5, 10, \ldots$
2. Compute the original EL statistic under the true $\mathcal{H}_0$ and a moment-based Bartlett correction
3. Obtain the bootstrap distribution of $-2 \log \mathrm{ELR}^*$ and compute both the bootstrap quantiles and the bootstrap Bartlett correction
   - Use three bootstrap variants: EL 'until the convex-hull condition is satisfied', ExEL1, and ExEL2
4. Test the null using these calibrations; report non-rejection rates (coverage probabilities)

1D simulation results (normal distribution)

Sample	$n = 5$				$n = 20$			
Nominal	.80	.90	.95	.99	.80	.90	.95	.99
EL	.66	.75	.81	.87	.77	.88	.93	.98
ExEL1	.66	.75	.81	.88	.77	.88	.93	.98
ExEL2	.66	.76	.82	.91	.77	.87	.93	.98
ExEL1-Bart-mom	.70	.79	.84	.90	.79	.89	.94	.98
ExEL1-Bart-boot	.78	.85	.89	.94	.79	.89	.94	.98
EL-boot-until	.71	.78	.85	.89	.79	.89	.95	.99
ExEL1-boot	.79	.89	.98	1.00	.79	.89	.95	.99

$n = 5$ [$n = 20$]: CH failure 6.5% [0%]; bootstrap CH failure 10.4% [0%]; extrapolation used in 32% [22%] of bootstrap replications.

2D simulation results (log-normal distribution)

Sample	$n = 10$				$n = 20$			
Nominal	.80	.90	.95	.99	.80	.90	.95	.99
EL	.33	.46	.56	.69	.40	.56	.68	.82
ExEL1	.33	.46	.56	.69	.40	.56	.68	.82
ExEL2	.33	.46	.56	.71	.40	.56	.67	.82
ExEL1-Bart-mom	.37	.50	.59	.72	.44	.59	.71	.84
ExEL1-Bart-boot	.59	.69	.76	.84	.54	.69	.78	.88
EL-boot-until	.63	.73	.79	.85	.74	.84	.90	.97
ExEL1-boot	.73	.86	.91	.98	.75	.85	.91	.97

$n = 10$ [$n = 20$]: CH failure 10% [0.4%]; bootstrap CH failure 12.9% [1.3%]; extrapolation used in 41% [32%] of bootstrap replications.

Simulation 2 design

$$Y = \beta_1 + \beta_2 \cdot X_1 + \beta_3 \cdot X_2^{\beta_4} + U, \quad \beta_0 = (0, 1, 1, 0.5)$$

Goal: estimate β_0 using the conditional moment restriction (CMR)

$$\mathbb{E}(\underbrace{Y - \beta_1 - \beta_2 X_1 - \beta_3 X_2^{\beta_4}}_{\rho(Z,\theta)} \mid X_1, X_2) = 0$$

Efficient estimation with CMRs requires *smoothing* and hence a choice of bandwidth b_n.

We choose an adaptive b_n such that each observation has exactly six neighbours.

Simulation 2 initialisations

1. Zero vector: $(0, 0, 0, 0)$
2. OLS under the mis-specified linear model with $\delta_0 = 1$
3. Smoothed Euclidean likelihood (SEuL) (also known as Conditional Euclidean Empirical Likelihood, a quadratic approximation to SEL)
4. Non-linear least squares (NLS), initialised at the OLS estimates
5. The true parameter $(0, 1, 1, 0.5)$ (oracle; infeasible in practice)

Simulation 2 results

Initialisation	Zero	OLS	SEuL	NLS	Truth
% at least one fail	99.99	52	32	54	50
Median iterations	34	29	24	31	32
% diverged	27	26	32	25	28

(1) At least one inner EL failure avoided by ExEL (so the outer optimiser can proceed);
(2) Median outer iterations of the BFGS optimiser;
(3) % of optimisations where any coordinate of $|\hat{\theta} - \theta_0|$ exceeds 10.

26% of all simulations had all five initialisations failing prior to extrapolation.

Showcase of `smoothemplik`

Practical implementation

New R package release on CRAN: `smoothemplik`!

- Empirical-likelihood methods for uni- and multi-variate data
- Non-parametric routines for kernel-based approaches
- EL variants to circumvent the convex-hull problem: AEL, BEL, ExEL
- Core functions in C++ for speed; many sanity checks and safe fall-backs

How to invoke ExEL

Let X be a matrix (e. g. the moment function $g(Z_i, \theta)_{i=1}^n$ evaluated at $\tilde{\theta}$). Let `mu` be the hypothesised column mean (e. g. $0, \ldots, 0$).

Make this simple call:

```
X <- g(Z, theta)  # Has q columns
EL(X, mu = mu, chull.fail = "taylor")  # ExEL1
EL(X, mu = mu, chull.fail = "wald")    # ExEL2
```

These methods will return the ELR statistic and the coefficients of the local parabola (ExEL1) / the exponential bridge (ExEL2).

Optional: observation weights are supported ⇒ smoothed empirical likelihood (SEL) is robust to the convex-hull-condition violation!

Method limitations and caveats

- ExEL does not yield individual probabilities $\hat{p}_i$ associated with observations – only the ELR statistic (i. e. what $\sum \log \hat{p}_i$ would have been)
- Extreme outliers can complicate tangency; back-offs mitigate but may need safeguards
- Level sets need not be convex because of the radial construction
- Choice of cuts affects locality; defaults work well in practice

Summary

So, why ExEL rather than ad-hoc fixes? Because ExEL:

- Preserves EL in the core region and supplies stable, concave, smooth quadratic tails elsewhere.
  - Interpretability is preserved; aids quasi-Newton optimisation
- Enables bootstrap on every data set and improves Bartlett-factor estimation via bootstrapping
- In terms of computational cost, remains comparable to baseline EL (only 2 or 3 times slower; ExEL1 can be as fast as EL)
- Gives accurate coverage at small n / heavy tails for the bootstrap and rescues many optimisation rounds with zero start
- Policy implications: small-sample inference becomes routine even when the convex-hull condition fails; this encourages wider use of EL-style methods in applied work.

Thank you for your attention and feedback!

CRAN:
smoothemplik

GitHub:
Fifis/smoothemplik

☆ 0 stars

Please star the project!

Solving the profiling problem

Solve the FOC for the Lagrangian:

$$\sum_{i=1}^{n} \log np_i - n\lambda \sum_{i=1}^{n} p_i(X_i - \theta) + \nu\Big(\sum_{i=1}^{n} p_i - 1\Big) = 0$$

$$\hat{p}_i = \frac{1}{n}\frac{1}{1+\hat{\lambda}'(X-\theta)}, \quad \hat{\lambda}\colon \sum_{i=1}^{n} \frac{X_i - \theta}{1+\hat{\lambda}'(X_i-\theta)} = 0$$

$\hat{\lambda}$ is found numerically via minimising the dual convex function

$$L(\lambda) := -\sum_{i=1}^{n} \log(1+\lambda'(X_i-\theta)).$$

Then $\hat{\lambda}(\theta)$ yields $\hat{p}_i(\hat{\lambda}(\theta))$ and hence $\mathrm{ELR}(\theta)$,

EL ratio test for hypotheses

For a p-dimensional parameter θ, the **log-EL ratio (ELR)** is defined as $\log \mathrm{ELR}(\theta) := \sum_i \log(1 + \hat\lambda(\theta)'(X_i - \theta))$.

Let $\hat\theta$ be the maximiser of $\log \mathrm{ELR}(\theta)$, and let θ_1 be the putative value. Then

$$-2[\log \mathrm{ELR}(\theta_1) - \log \mathrm{ELR}(\hat\theta)] \sim \chi^2_{\dim \theta_1}$$

(For testing hypotheses on selected parameters, treat the rest as nuisance parameters and profile them out.)

Types of EL calibration

- χ^2 calibration: using asymptotic critical values $Q_{\chi^2_p}(1-\alpha)$ gives overly narrow confidence regions
- Critical values from $F_{\dim \theta_0, n-\dim \theta_0}$: a conservative improvement
- Bartlett correction: dividing $-2\widehat{\log \mathrm{ELR}}(\theta_0)$ by $(1+b/n)$ gives $O(n^{-2})$ coverage error; choosing b is not always easy
- Bootstrap provides excellent approximation with $O(n^{-2})$ error but is computationally heavy
- Adjusted EL/Balanced EL: add 1–2 specially chosen points to the dataset; simpler than bootstrap, gives $O(n^{-2})$ coverage error

How often does the CH condition fail?

Failure probability in certain cases has a convenient closed-form expression.

- In one dimension, the probability of failure is $F(\theta_0)^n + (1 - F(\theta_0))^n$
- With independent coordinates, the axis-aligned approximation is $1 - \prod_j [1 - F_j(\theta_0^{(j)})^n - (1 - F_j(\theta_0^{(j)}))^n]$
- Centrally symmetric laws: failure rate increases with d, decreases with n

In higher dimensions, failures are more likely (Wendel's theorem).

Partial solution: self-concordant EL

If θ lies outside the convex hull of the data, then, there are no such positive p_i: $\sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i X_i = \theta$.

For λ search, consider replacing the logarithm with its 4th-order Taylor expansion below $1/n$ (Owen, 2013, CJS):

$$\log^* x := \begin{cases} \log x, & x \geq \frac{1}{n}, \\ \log \frac{1}{n} - \frac{25}{12} + 4nx - 3(nx)^2 + \frac{4}{3}(nx)^3 - \frac{1}{4}(nx)^4, & x < \frac{1}{n} \end{cases}$$

Then, $-\log^*(1 + \lambda'(X_i - \theta))$ is always finite $\Rightarrow$ all initial values may work, and Newton search with backtracking converges if θ is in the convex hull, but never terminates otherwise.

ExEL2 implementation

There are two possible scenarios: $\log \text{ELR}(\theta) > -0.5W(\theta)$ (transition must occur to a lower function) and $\log \text{ELR}(\theta) \le -0.5W(\theta)$ (transition to a higher function). They require different kinds of bridges.

Bridge 1: $\alpha_0 + \alpha_1\theta + \alpha_2 \exp(\theta - \theta_{\max})$.

Bridge 2: $\beta_0 + \beta_1\theta + \beta_2 \exp(-(\theta - \theta_{\max}))$

In the appendix, I provide the necessary derivations: the objective to find the root of and the expressions for α's and β's as a function of the root.

Bartlett correction

The correction to $-2\log \mathrm{ELR}$ can be multiplicative (DiCiccio, Hall, Romano, 1991, *Ann. Stat.*) – the **Bartlett correction**:

$$\mathbb{P}(-2\log \mathrm{ELR}(\theta_0) \le t) = \mathbb{P}(\chi^2_p \le t) + O(n^{-1})$$

$$\mathbb{P}(-2\log \mathrm{ELR}(\theta_0)[1 - b/n] \le t) = \mathbb{P}(\chi^2_p \le t) + O(n^{-2})$$

b can be estimated with long analytic formulæ or via the bootstrap:

$$\widehat{1 - b/n} = \overline{-2\log \mathrm{ELR}^{*(b)}}/p$$

In practice, trimmed mean works better in extremely small samples.

1D simulation results (log-normal distribution)

Sample	$n = 5$				$n = 20$			
Nominal	.80	.90	.95	.99	.80	.90	.95	.99
EL	.56	.65	.70	.76	.70	.80	.87	.93
ExEL1	.56	.65	.71	.78	.70	.80	.87	.93
ExEL2	.56	.65	.71	.78	.70	.80	.86	.91
ExEL1-MBart	.62	.69	.74	.80	.75	.84	.89	.94
ExEL1-BBart	.70	.76	.80	.85	.75	.84	.89	.94
EL-boot-until	.58	.66	.73	.78	.75	.86	.91	.96
ExEL1-boot	.71	.79	.92	.97	.75	.86	.91	.96

$n = 5$ [$n = 20$]: CH failure 16.0% [0.0%]; bootstrap CH failure 14.5% [0.3%]; extrapolation used in 33% [26%] of bootstrap replications.

2D simulation results (normal distribution)

Sample	$n = 10$				$n = 20$			
Nominal	.80	.90	.95	.99	.80	.90	.95	.99
EL	.46	.61	.71	.84	.51	.69	.80	.93
ExEL1	.46	.61	.71	.84	.51	.69	.80	.93
ExEL2	.46	.61	.71	.86	.51	.69	.80	.93
ExEL1-MBart	.50	.65	.75	.87	.53	.71	.82	.94
ExEL1-BBart	.61	.75	.83	.92	.55	.73	.83	.94
EL-boot-until	.76	.87	.92	.96	.79	.89	.95	.99
ExEL1-boot	.81	.92	.97	1.00	.79	.89	.95	.99

$n = 10$ [$n = 20$]: CH failure 2.0% [0.0%]; bootstrap CH failure 5.0% [0.0%]; extrapolation used in 34% [25%] of bootstrap replications.

Simulation 2 median estimates

Initialisation	Zero	OLS	SEuL	NLS	Truth
$\hat{\alpha}\,(\alpha_0 = 0)$	-0.15	-0.12	-0.31	-0.10	-0.22
$\hat{\beta}\,(\beta_0 = 1)$	1.00	1.00	1.00	1.01	1.00
$\hat{\gamma}\,(\gamma_0 = 1)$	1.13	1.05	1.29	1.03	1.19
$\hat{\delta}\,(\delta_0 = 0.5)$	0.42	0.57	0.43	0.59	0.42