FACULTY OF LAW, ECONOMICS AND FINANCE

Department of Economics and Management

Discussion Paper

2025-19

Economics

Department of Economics and Management
University of Luxembourg

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

https://www.uni.lu/fdef-en/research-departments/department-of-economics-and-management/publications/

Andreï Kostyrka, DEM, Université du Luxembourg, UL

Dec. 2025

For editorial correspondence, please contact: dem@uni.lu
University of Luxembourg
Faculty of Law, Economics and Finance
4, rue Alphonse Weicker
L-2721 Luxembourg

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

Andreï Victorovitch KOSTYRKA
Department of Economics and Management, University of Luxembourg*

16th December 2025

Abstract

Empirical likelihood (EL) breaks down when the hypothesised mean falls outside the convex hull of the sample. We propose extrapolated EL (ExEL) – two splicing schemes that extend the log-EL ratio beyond the hull while leaving it unchanged on a user-chosen interior region. The first scheme, ExEL1, continues EL past a data-driven cut-off using its local quadratic (Taylor) expansion. The second scheme, ExEL2, smoothly splices EL to its global Wald quadratic approximation via a convex bridge. Both methods extend naturally to multiple dimensions by radial reduction. In simulations with small samples – where convex-hull violations are common – ExEL remains well-behaved and distinguishes mild from severe violations. It also has attractive inferential properties, delivering accurate coverage probabilities with bootstrap calibration.

1 Introduction

Empirical likelihood (EL; Owen, 1988, 2001) is a popular non-parametric likelihood framework for estimation and inference under moment restrictions. Many moment-condition models in economics, biology, physics, and other sciences can be estimated via EL maximising the profile empirical likelihood over $\theta \in \Theta \subset \mathbb{R}^p$. Computation typically proceeds in two loops: an inner maximisation over weights (for a fixed θ subject to the empirical estimating equations) and an outer search over θ (for confidence regions or tests). Equivalently, EL re-weights the empirical distribution so that selected functionals match specified values of θ.

A notorious impediment to EL-based modelling is the convex-hull constraint: the inner optimisation problem is feasible if and only if the zero vector of moments lies in the convex hull of the transformed sample. When this holds, then there exists a re-weighted empirical distribution whose functionals satisfy the constraints. When it fails, the profile log-EL is undefined (conventionally $-\infty$), rendering mild and gross violations of parametric hypotheses about the true parameter value θ_0 indistinguishable.

To deal with this problem, we propose *extrapolated EL* (ExEL): two splicing / extrapolation schemes that (1) coincide with EL on a large-enough user-chosen set, (2) extend it smoothly and monotonically beyond the hull, and (3) deliver Wald-type tails to enable standard χ^2 calibration. Our contributions are threefold. First, we provide a local Taylor splice (ExEL1) that replaces EL by its quadratic expansion at a data-driven cut point; it is computationally simple and preserves smoothness. Second, we propose a Wald splice (ExEL2) that 'glues' EL to the appropriate Wald parabola using either a common supporting line (when it exists) or a strictly convex exponential

*We thank Gautam Tripathi, Benjamin Holcblat, and Ali Atabaigialami for their helpful comments. We also thank seminar participants at the University of Luxembourg for their insightful suggestions.

bridge, ensuring convexity and valid Wald asymptotics outside the hull. Third, we give a multivariate reduction that uses radial casting profiles along rays, avoiding any explicit hull computation. We study coverage accuracy in designs where convex-hull violations are non-negligible. The proposed splices are drop-in replacements for the profile EL in optimisation routines that cannot tolerate non-finite objective values.

A key virtue of ExEL is that it coincides with the classical EL on practically large regions where the deviations from the null hypothesis are modest. All well-known interior refinements therefore apply without modification: Bartlett correction (reducing coverage error from $O(n^{-1})$ to $O(n^{-2})$ DiCiccio et al. (1991); adjusted EL (AEL), which adds a small amount of pseudo-mass to improve small-sample behaviour (J. Chen et al., 2008); balanced EL (BEL), which adds two points (Emerson & Owen, 2009); and extended EL (EEL), which maps outside the hull via a similarity transform (Tsao, 2013; Tsao & Wu, 2013, 2014). Because ExEL coincides with EL on the feasible set, any Bartlett or other interior calibration chosen for EL carries over verbatim there.

1.1 Literature review

The convex-hull issue is a particular case of the empty-set problem (Bergsma et al., 2012; Grendár & Judge, 2009), where the set of distributions meeting the moment conditions may be empty for all values of θ. In empirical estimating equations, replacing the model with the empirical distribution makes feasibility equivalent to the origin lying in the convex hull of the transformed data; there may be no θ for which this holds. Remedies fall broadly into four classes.

Euclidean likelihood (Antoine et al., 2007; Brown & Chen, 1998; Owen, 2001) and, more generally, Cressie–Read discrepancies (Baggerly, 1998; Smith, 2007) relax non-negativity of implied probabilities, allowing negative weights when the convex-hull condition fails. These methods can work well for estimation, though their interpretation differs from EL's likelihood-based one.

Bartolucci (2007) adds a quadratic penalty to the profile EL and relaxes the moment constraint, extending the domain to the full parameter space. In doing so, he ensures that the penalised EL ratio is always finite, and if the slackness parameter is of order $O(1/\sqrt{n})$, the penalised EL ratio converges in distribution to the original EL. Practical performance depends on penalty choice and calibration. Lahiri and Mukhopadhyay (2012) adapt this method to $p > n$ problems, and Zhang and Shao (2016) use it for block-wise EL to bound coverage probabilities under weak dependence.

Adjusted EL (AEL) (J. Chen et al., 2008) augments the sample with one or more artificial points proportional to the negative sample mean, guaranteeing feasibility for all θ. Y. Liu and Chen (2010) show that AEL is Bartlett-correctable and propose adding two artificial observations; J. Chen and Huang (2013) show that AEL performs well in small samples in terms of coverage. Balanced AEL (BAEL) (Emerson & Owen, 2009; Nguyen et al., 2015) adds a second point so that the original and augmented sample means coincide, improving finite-sample behaviour. Baragona et al. (2015) notice that AEL-based confidence sets may sometimes be as large as the full parameter space.

Tsao (2013) and Tsao and Wu (2013, 2014) map EL contours to a larger set with the same centre, ensuring finiteness for all θ (extended EL). Consistency is ensured if the expansion factor approaches one from above as $n \to \infty$. This preserves confidence-set boundary shapes and often attains coverage competitive with Bartlett-corrected EL. Our radial reduction does not preserve exact similarity of regions but has the benefit of retaining the original EL in a user-chosen interior.

For broader surveys of convex-hull remedies, see Baragona et al. (2017) and, more generally, P. Liu and Zhao (2022).

While our exposition focuses on the mean, the convex-hull pathology and our splices apply to EL for general estimating equations (GEE) $\mathbb{E}g(Z,\theta) = 0$ with smooth g (Qin & Lawless, 1994). Here, the profile EL re-weights the empirical distribution subject to $\sum_i p_i g(Z_i,\theta) = 0$, and infeasibility occurs precisely when the origin lies outside the convex hull of $g(Z_i,\theta)$. The broader generalised empirical likelihood (GEL) family (Newey & Smith, 2004) replaces the log with other convex generators (exponential tilting, Euclidean, Cressie–Read), but the domain mismatch persists whenever feasibility

is enforced via non-negative weights. Our construction thus provides a generic outer-domain completion for smooth profile objectives with EL-type local quadratic structure and explosive behaviour near hull boundaries.

The article is organised as follows. Section 2 formulates the convex-hull problem in the context of EL and provides real-world examples. Section 3 proposes two one-dimensional solutions. Section 4 extends them to the multi-variate setting. Section 5 reports numerical comparisons with existing methods. Section 6 concludes.

2 Empirical likelihood and the convex-hull problem

2.1 Empirical likelihood and the spanning condition

Let $Z_1, \dots, Z_n \in \mathbb{R}^d$ be IID draws from an unknown distribution F. The non-parametric maximum likelihood estimator of $F(z)$ is given by $\hat{F}(z) := \sum_{i=1}^n \hat{p}_i \mathbb{I}(Z_i \le z)$, where $\hat{p}_1, \dots, \hat{p}_n$ maximise the empirical likelihood,

$$L(p_1, \dots, p_n) := \prod_{i=1}^n p_i,$$

or, equivalently, its logarithm, $\ell := \log L$.

Now consider a model defined by q unconditional moment restrictions that identify the p-dimensional parameter θ_0 if they hold under F:[1]

$$\theta_0 \colon \mathbb{E} g(Z, \theta_0) = \int_{\mathbb{R}^d} g(z, \theta_0) \, \mathrm{d}F(z) = 0_{q \times 1}. \tag{2.1}$$

θ_0 is locally identified if the Jacobian $\frac{\partial}{\partial \theta} \mathbb{E} g(Z, \theta_0)$ has full column rank. This condition is enough for global identification of θ_0 if $\theta \mapsto g(Z, \theta)$ is linear (Newey & McFadden, 1994, Section 2.2.3).

The *profile EL* maximises the inner objective under the moment and simplex constraints for a fixed θ:

$$\ell(\theta) := \max_{p_1, \dots, p_n > 0} \sum_{i=1}^n \log p_i \quad \text{s. t.} \quad \sum_{i=1}^n p_i = 1, \quad \sum_{i=1}^n p_i g(Z_i, \theta) = 0_{q \times 1}. \tag{2.2}$$

The solution to (2.2) is

$$\hat{p}_i(\theta) = \frac{1}{n} \frac{1}{u_i(\theta)}, \quad u_i(\theta) := 1 + \hat{\lambda}(\theta)' g(Z_i, \theta), \quad \hat{\lambda}(\theta) \colon \sum_{i=1}^n \frac{g(Z_i, \theta)}{u_i(\theta)} = 0_{q \times 1},$$

and the outer maximisation

$$\hat{\theta}_{\mathrm{EL}} := \arg\max_{\theta \in \Theta} \ell(\theta)$$

yields the EL estimator.[2]

Define the empirical likelihood ratio (ELR) and double the negative of its logarithm (NLELR):

$$\mathrm{ELR}(\theta) := \frac{L(\theta)}{L(\tilde{p}_1, \dots, \tilde{p}_n)} = \prod_{i=1}^n n \hat{p}_i(\theta), \quad \mathrm{NLELR}(\theta) := -2 \log \mathrm{ELR}(\theta).$$

In the just-identified case, under $\mathscr{H}_0 \colon \theta_0 = c$ and standard regularity conditions (Owen, 2001, Theorem 3.2), $\mathrm{NLELR}(\theta_0) \xrightarrow[n \to \infty]{d} \chi^2_p$. In the over-identified case, Qin and Lawless (1994, Theorem 2) state that for testing $\mathscr{H}_0$, the profile EL ratio statistic $\mathrm{NLELR}(\theta_0) - \mathrm{NLELR}(\hat{\theta}_{\mathrm{EL}}) \xrightarrow[n \to \infty]{d} \chi^2_p$ can be

[1] The case of conditional moment restrictions is studied by Ai and Chen (2003), Kitamura et al. (2004) and Smith (2007) and can be handled, e. g., by introducing n one-dimensional local unconditional restrictions.

[2] u_i depends on $\hat{\lambda}$, but this dependence is suppressed for notational convenience.

used, while the over-identification statistic for testing (2.1) is $\mathrm{NLELR}(\hat{\theta}_{\mathrm{EL}}) \xrightarrow[n\to\infty]{d} \chi^2_{q-p}$. Hence, a $(1-\alpha)$-confidence region for θ_0 that is valid for $q \geq p$ is $\{\theta\colon \mathrm{NLELR}(\theta) - \mathrm{NLELR}(\hat{\theta}_{\mathrm{EL}}) \leq Q_{\chi^2_p}(1-\alpha)\}$.

A convenient approximation of the EL ratio is the Wald quadratic. A second-order expansion of ℓ at $\hat{\theta}_{\mathrm{EL}}$ gives

$$\ell(\theta) \approx \ell(\hat{\theta}_{\mathrm{EL}}) + \nabla\ell(\hat{\theta}_{\mathrm{EL}})(\theta - \hat{\theta}_{\mathrm{EL}}) + \frac{1}{2}(\theta - \hat{\theta}_{\mathrm{EL}})\nabla^2\ell(\hat{\theta}_{\mathrm{EL}})(\theta - \hat{\theta}_{\mathrm{EL}}).$$

$\nabla\ell(\hat{\theta}_{\mathrm{EL}}) = 0_{p\times 1}$ (first-order condition for $\hat{\theta}_{\mathrm{EL}}$), and the EL ratio statistic is approximated by the Wald statistic

$$2[\ell(\theta) - \ell(\hat{\theta}_{\mathrm{EL}})] \approx W(\theta) := (\theta - \hat{\theta}_{\mathrm{EL}})'[\nabla^2\ell(\hat{\theta}_{\mathrm{EL}})](\theta - \hat{\theta}_{\mathrm{EL}}), \tag{2.3}$$

where the Hessian $\nabla^2\ell$ is positive definite for large-enough n.

The Lagrange multiplier admits the dual characterisation

$$\hat{\lambda}(\theta) := \operatorname*{arg\,min}_{\lambda\in\mathbb{R}^q} -\sum_{i=1}^{n} \log(1 + \lambda' g(Z_i, \theta)), \tag{2.4}$$

where the objective is strictly convex on $\{u_i(\theta) > 0\ \forall i\}$. Hence one may solve for $\hat{\lambda}$ either by root-finding or convex optimisation.[3] Following Kitamura (2007, Section 8.1), we refer to the condition $0 \in \mathrm{Conv}\{g(Z_i,\theta)\}_{i=1}^n$ as the *spanning condition* because finding $\hat{\lambda}$ is an ill-posed problem when zero is not *spanned* by the columns of $\{g(Z_i,\theta)\}$. When spanning fails, $\ell(\theta)$ is defined to be $-\infty$, and $\mathrm{NLELR}(\theta)$ is taken as $+\infty$. In high dimensions, when convex-hull algorithms become computationally prohibitive, the spanning-condition failure can be detected indirectly when a line-search-based λ solver stalls with a very large $\|\hat{\lambda}\|$.

Let $\Delta_n := \{\{p_i\}_{i=1}^n\colon p_i > 0\ \forall i,\ \sum_{i=1}^n p_i = 1\}$ be the simplex and let $S_\Delta(\theta) := \{\{p_i\} \in \Delta_n\colon \sum_{i=1}^n p_i\, g(Z_i,\theta) = 0_{q\times 1}\}$ be the set of weights satisfying the simplex and the moment constraints. Define the parameter *feasible set* as $S_\Theta := \{\theta \in \Theta\colon S_\Delta(\theta) \neq \varnothing\}$; $\theta \in S_\Theta$ implies the existence of a valid solution to (2.2), i. e. $\ell(\theta)$ is well-defined. Equivalently, $S_\Theta = \{\theta \in \Theta\colon 0 \in \mathrm{Conv}\{g(Z_i,\theta)\}_{i=1}^n\}$. As θ approaches the feasible-set boundary, $\mathrm{NLELR}(\theta) \to \infty$, i. e. it diverges at the hull because at least one $u_i(\theta) \to 0$. This is numerically problematic because many optimisers require finite objective values. Quasi-Newton optimisers like BFGS usually start optimisation w. r. t. θ by finding a search direction and choosing an acceptable step along that direction. However, a trial step may violate spanning; some implementations recover by backtracking, others do not. When no good initial value is available, even the starting point may be infeasible. Constraining the optimiser to the region where the spanning condition holds is not feasible when the number of regressors exceeds 8–10 (this includes the majority of empirical applications in econometrics) because convex-hull computation scales poorly.[4] Moreover, infeasibility at a candidate point is itself strong evidence against the hypothesis and ought to be reflected in the objective, not merely treated as a non-finite value.

To obtain directional guidance on the full parameter space, we require a surrogate that penalises distance from the feasible set and therefore propose two variants of *extrapolated empirical likelihood* (ExEL): a completion of NLELR to $\mathbb{R}^p$ that agrees with EL on a large part of its natural domain and is smooth, monotone, and convex outside. It uses analytical derivatives and, crucially, distinguishes 'mild' violations from 'gross' violations of parametric hypotheses about θ_0. A strength of ExEL is that it preserves the EL shape – and hence confidence regions – on a user-chosen admissible region, unlike AEL (J. Chen et al., 2008; Y. Liu & Chen, 2010) or BAEL (Emerson & Owen, 2009), which modify the objective everywhere. Thus, within the feasible set, all standard properties (e. g. Bartlett correctability) are retained. When the spanning condition fails in some direction, the extrapolated

[3] If $q = 1$, then, the box constraints $p_i < 1\ \forall i$ put a convenient range restriction on λ for uni-variate root search.

[4] Worst-case complexity of the Quickhull algorithm in d dimensions is $O(n^{\lfloor d/2\rfloor})$; e. g. computing a 16-dimensional convex hull of 1000 points may involve processing $\approx 10^{24}$ facets.

statistic provides a reasonable (and well-calibrated for the ExEL2 variant) penalty reflecting the distance to the data-supported set.

We develop two complementary constructions. ExEL1 replaces logELR beyond a data-driven cut point by its quadratic (Taylor) polynomial, ensuring monotonicity, $\mathscr{C}^1$ continuity, and strict convexity of the extrapolated branch. ExEL2 splices logELR to the directional Wald parabola: a strictly concave affine-exponential bridge provides a $\mathscr{C}^1$ transition to the Wald tail. Our approach differs from the element-wise Taylor devices of Owen (2001) and Owen (2013), which extrapolate $\log_* u_i(\theta)$ term-by-term; we extrapolate NLELR(θ) directly because it is the profile likelihood that practitioners often use in optimisation and inference. Moreover, extrapolated $\log_* u_i(\theta)$ does not lead to finite $\hat{\lambda}$ and therefore logELR values in case of a convex-hull violation; our extrapolated construction foregoes the $\hat{\lambda}$ search altogether.

For clarity we present right-hand (monotone-increasing) extrapolation; the left-hand case follows by symmetry.

2.2 Why spanning-condition failure is a pervasive problem

Failure of the convex hull to contain the true target is a *finite-sample issue*. In one dimension, with $\mathbb{E}Z = \theta_0$ and $Z_1, \ldots, Z_n$ are IID, the failure probability is the probability that all observations are lying to one side of the true mean, i. e. $F(\theta_0)^n + (1 - F(\theta_0))^n$. In p dimensions, for centrally symmetric continuous distributions, Wendel's theorem gives $\Pr(\text{CH failure}) = 2^{1-n} \sum_{k=0}^{p-1} \binom{n-1}{k}$, which increases with p (it is harder to 'surround' the mean in higher dimensions) and decreases with n. Without central symmetry but with independent coordinates, an axis-aligned lower bound is $\Pr(\text{CH failure}) \geq 1 - \prod_{j=1}^{p} \left[1 - F_j(\theta_0^{(j)})^n - (1 - F_j(\theta_0^{(j)}))^n\right]$, which likewise increases with p and decreases in n. More general bounds are given by Tsao (2004), e. g. if $p \to \infty$, then EL ratios may degenerate into a point mass in high-dimensional models; consequently, for $p \gg n$, high confidence levels may be unattainable.

These failures arise routinely in econometric applications where the spanning condition is easy to violate.

1. Bootstrap failure. In small samples, bootstrap calibration often improves EL because raw NLELR can exceed χ^2 quantiles (Emerson & Owen, 2009, Section 5.1). However, for the bootstrap to work, the spanning condition must hold in every replication; when the support is impoverished by $\approx$36.8% failures, a bootstrap scheme can deliver coverage probabilities only up to $1 - \Pr(\text{NLELR} = \infty)$ (Y. Liu & Chen, 2010, Section 2.2). Owen (2001, Section 3.3) reports 23/1000 samples with the mean outside the 4D hull. He notes that convex-hull failure is more likely for skewed data and concludes that this restriction can be a critical diagnostic if the fraction of bootstrap samples with such a configuration exceeds a certain level. Therefore, bootstrap in small samples – where it is the most desirable for calibration – is more likely to fail with unadjusted EL.

2. Optimiser failure. Modern solvers probe many θ; whenever a probe violates the spanning condition, EL returns $-\infty$ / `NaN` / `NA`. Some routines (e. g. L-BFGS-B in **R**) require finite objectives and may exit once non-finite values or undefined gradients appear.

E. g., for non-linear regression via EL, the estimating function $g(Z, \theta)$ is $\nabla_\theta f(Z_i, \theta)(Y_i - f(X_i, \theta))$. There is no general formula for the initial value to start numerical search in the non-linear case; a common choice is $\theta_0 = \vec{0}$. However, the residuals at the zero initialisation may have the same sign for all observations. If $0 \notin \text{Conv}\{g(Z_i, \theta_0)\}_{i=1}^n$, moment constraints cannot be solved with positive affine weights, and $\text{logELR}(\theta_0) = -\infty$. This is a feasibility (geometry) issue, not a calibration issue.

3. Many dummy variables. Consider a linear regression $Y = \alpha_0 + X'\beta_0 + D'\gamma_0 + U$ with numerous indicators D (units, families, regions etc.). If during optimisation for a tentative $\tilde{\theta}$ in some group all residuals share the same sign, the corresponding weighted moments cannot span zero, and optimisation cannot proceed.

E. g., this problem is especially serious in smoothed empirical likelihood (SEL) (Kitamura et al., 2004) with sparse cells. Consider the SEL estimator of a linear model with many indicator regressors,

where the data are split into numerous small cells defined by interactions of dummy variables. Some cells may have very few points; within such cells, the origin is easily excluded from the local hull of the moment-function values during optimisation. Smoothing on discrete variables may or may not partially mitigate the problem through coarsening because it may not alter the hull geometry.

These considerations motivate a surrogate that reconstructs pseudo-values in regions where EL is $-\infty$ because of convex-hull violations. To our knowledge, this is the first approach that completes the profile log-EL ratio directly: it preserves the original EL on a chosen middle region and extrapolates it to $\mathbb{R}^p$, enabling standard calibration methods (bootstrap, Bartlett etc.).

Our approach does not yield individual implied probabilities $\hat{p}_i$ for the original observations – it extrapolates $\ell(\theta)$ and projects what $\sum_i \log \hat{p}_i$ would be approximately equal to under no violation. Many existing methods rely on adding pseudo-observations – typically 1 or 2 – to the sample and maximising the EL with $n+1$ or $n+2$ observations. However, neither approximating the log-likelihood nor adding surrogate observations provides an incontrovertible probabilistic interpretation of the modified EL. With AEL and BEL, the implied probabilities do not correspond to the original empirical distribution (the sum of probabilities on the original n points is less than 1). In addition, interpretation of pseudo-observations may be hindered by the fact that the moment conditions may correspond to positive quantities, e. g. incomes, heights, populations, distances etc. – in which case probabilities of negative amounts have no obvious interpretation. In this sense, when ExEL1/ExEL2 is invoked with no convex-hull-condition violation at the optimum (and NLELR is not exploding), it retains the probabilistic interpretation of the original EL because it coincides with EL, unlike AEL/BEL. When the convex-hull condition fails, neither ExEL1/2 nor AEL/BEL can directly answer the question 'what are the $\hat{p}_i > 0$ on the original n observations such that the moment conditions hold'.

3 Extrapolated empirical likelihood (ExEL)

In this section, we show how to extrapolate EL for the mean with $g(Z,\theta) = Z - \theta$ and $q = p = 1$ to simplify exposition. A multi-variate extension of the proposed technique to general estimating equations is provided in the next section (under standard regularity conditions; see, e. g., S. X. Chen and Cui, 2007, Section 2.1). To simplify notation, we write f for the target profile (NLELR). We focus on the right branch of $\theta \mapsto \mathrm{NLELR}(\theta)$; the left one is handled symmetrically.

3.1 Taylor extrapolation outside the convex hull (ExEL1)

The following mathematical construction is valid for any convex function f, but in the EL application, we are using $f(\theta) = \mathrm{NLELR}(\theta)$. Hence, consider any convex $f \in \mathscr{C}^2(a,b)$ for which $\lim_{\theta\to\theta_{\max}} f^{(k)}$ exists for $k \in \{0,1,2\}$, where $\theta_{\max} \in S_\Theta$ is the desired splice location. For the uni-variate mean EL, f is strictly convex on (a,b), where a and b denote the minimum and the maximum of the sample, respectively (proof in Appendix A.1). Therefore, f is uni-modal with a unique minimum $\hat{\theta}$ (in this case, the sample average $\bar{Z}_n$). Let $\theta_{\max}$ satisfy $\hat{\theta} < \theta_{\max} < b$ and $f(\theta_{\max}) =: f_n^{\max}$, where $f_n^{\max}$ is a value chosen by the researcher taking into account the sample size (the choice of the cut-off level is explained in Section 3.3).

An extrapolation algorithm for the right branch that locates $\theta_{\max}$ and replaces f smoothly by its second-order Taylor expansion for $\theta > \theta_{\max}$ is:

1. Choose the maximum level of f, $f_n^{\max}$, beyond which it is extrapolated, e. g. the $(1 - 1/n^{1+\delta})$ quantile of χ^2_1 for some $\delta > 0$;
2. Use uni-variate root search (e. g. Brent's method) to find $\theta_{\max} > \hat{\theta}$ such that $f(\theta_{\max}) = f_n^{\max}$. Because f is convex, uni-modal, and non-decreasing on $[\hat{\theta}, b)$, the right-branch solution

is unique. In practice, bracket with $\hat{\theta}$ fixed as the lower end-point and expand the right end-point until f crosses $f_n^{\max}$ upwards.

3. Compute $f(\theta_{\max})$ and obtain $f'(\theta_{\max})$ and $f''(\theta_{\max})$ (analytically or numerically). For EL, the analytic formulæ are given in Appendices A.1 and D.
4. Form the quadratic Taylor polynomial about $\theta_{\max}$, $a_1\theta^2 + b_1\theta + c_1$ with $a_1 = 0.5f''(\theta_{\max})$, $b_1 = f'(\theta_{\max}) - f''(\theta_{\max})\theta_{\max}$, $c_1 = f(\theta_{\max}) - f'(\theta_{\max})\theta_{\max} + 0.5f''(\theta_{\max})\theta_{\max}^2$.

If $f \in \mathscr{C}^3$ near $\theta_{\max}$, the approximation error has remainder $\frac{1}{6}f'''(\xi)(\theta-\theta_{\max})^3$ for some $\xi \in (\theta_{\max}, \theta)$. Even when only $\mathscr{C}^2$ holds, the quadratic splice is a stable finite surrogate of $f(\theta)$ far from $\hat{\theta}$ and is therefore useful in optimisation owing to its finiteness. Splicing both ends (with $\theta_{\min}$ the left-branch splice point) yields

$$f_*(\theta) := \begin{cases} a_0\theta^2 + b_0\theta + c_0, & \theta < \theta_{\min}, \\ f(\theta), & \theta_{\min} \le \theta \le \theta_{\max}, \\ a_1\theta^2 + b_1\theta + c_1, & \theta > \theta_{\max}, \end{cases}$$

which is $\mathscr{C}^1$ by construction: $\lim_{\theta\to\theta_{\max}^-} f_*(\theta) = \lim_{\theta\to\theta_{\max}^+} f_*(\theta)$ and $\lim_{\theta\to\theta_{\max}^-} f_*'(\theta) = \lim_{\theta\to\theta_{\max}^+} f_*'(\theta)$ (and similarly at $\theta_{\min}$). Applied to the profile EL objective, this gives the extrapolated empirical likelihood of the first type (ExEL1).

Avoiding the root solver. In uni-variate mean EL, the boundary of the convex hull is the two outermost observations. Root search for $\theta_{\max}$ may require many evaluations and can be bypassed if $\theta_{\max}$ is re-defined as the penultimate order statistic, $Z_{(n-1)}$, or even $(Z_{(n-1)} + Z_{(n)})/2$. If $Z_{(n-1)}$ is too close to $Z_{(n)}$, then the slope $f'(Z_{(n-1)})$ might be too steep, and numerical differentiation may be unreliable; then choose a different $\theta_{\max}$ (e. g. $Z_{(n)}$ minus the median gap between $Z_{(i)}$) to avoid numerical blow-up near $Z_{(n-1)}$ while keeping the correction asymptotically negligible. This loses the option to set $f_n^{\max}$ directly, but in practice defining $f_n^{\max}$ through $\theta_{\max}$ does not affect accuracy in optimisation and bootstrapping problems, while the speed gains are substantial.

Whilst this extrapolation is easy to implement and to use for estimation, it is not intended for extreme hypothesis testing when the hypothesised mean lies beyond $\theta_{\max}$. The extrapolated branch is generally not equal to the directional Wald parabola (centred at the unconstrained estimate) in (2.3) and therefore does not correspond to the tail of an approximately χ_1^2 random variable. However, for parameter estimation in moment-based models, this approach is attractive for optimisers that cannot accept non-finite function values and therefore explore the exterior of the convex hull. An objective that is quadratic almost everywhere (except on $[\theta_{\min}, \theta_{\max}]$) makes numerical optimisation routines perform remarkably well. If analytical derivatives of f are not available, a well-programmed implementation of ExEL1 should use memoisation to store pre-computed coefficients (a_0, b_0, c_0) and (a_1, b_1, c_1) to speed up the line-search step of the BFGS algorithm and enable instant visual diagnostics of f.

3.2 Transition into Wald statistic outside the convex hull (ExEL2)

A second way to extend f is to splice it smoothly into a pre-specified parabola on each side. For NLELR, this parabola is the Wald statistic in (2.3). This approach requires more computation, but it has a clear inferential gain: far from the sample mean, the splice inherits the familiar χ^2-type tail behaviour of Wald tests (notwithstanding that Wald confidence intervals themselves may be unreliable for extreme hypotheses; cf. Cosma et al. (2026, Section 4.3)).

Let $\bar{Z}_n$ be the sample average of the $Z_1, \ldots, Z_n$ and $V_{\bar{Z}} = \widehat{\mathrm{Var}}\,\bar{Z}_n = \frac{1}{n}\big(\frac{1}{n}\sum_{i=1}^n (Z_i - \bar{Z}_n)^2\big)$ the estimated variance of the sample average. If the data have weights, compute the weighted average and weighted variance; see Appendix D. The Wald parabola for testing $\mathscr{H}_0 : \mathbb{E}Z = \theta$ is the squared t-statistic $(\theta - \bar{Z}_n)^2 / V_{\bar{Z}}$. For simplicity, *centre the data* (by subtracting $\bar{Z}_n$) and write $W(\theta) := a\theta^2$ with $a = 1/V_{\bar{Z}}$.

Since it is not always possible to join $f(\theta)$ and $W(\theta)$ with a straight line while preserving convexity and smoothness, we interpolate to $W(\theta)$ over a buffer interval using a strictly convex 'bridge'.[5] Unlike ExEL1, there are two possible constructions. If for a given $\theta_1 > 0$, $f(\theta_1) > W(\theta_1)$, then the bridge $h_1(\theta)$ should be nearly linear at θ_1 (tangent to f) and then grow rapidly to touch W at some $\theta_2 > \theta_1$ (derivation in Appendix B.1):

$$h_1(\theta) := a_0 + a_1\theta + a_2 \exp(\theta - \theta_1).$$

If $f(\theta_1) < W(\theta_1)$, the bridge $h_2(\theta)$ should touch f in a shallow, U-shaped fashion at θ_1 and be nearly linear at $\theta_2 > \theta_1$ (derivation in Appendix B.2):

$$h_2(\theta) := a_0 + a_1\theta + a_2 \exp(-(\theta - \theta_1)).$$

Define

$$f_{**}(\theta) = \begin{cases} f(\theta), & \theta < \theta_1, \\ h_1(\theta) \text{ or } h_2(\theta), & \theta_1 \le \theta \le \theta_2, \\ W(\theta), & \theta > \theta_2, \end{cases}$$

which is $\mathscr{C}^1$, convex, and strictly convex on (θ_1, θ_2) once the parameters are chosen as follows.

Choose any $\theta_1 > 0$ and check whether $f(\theta_1) > W(\theta_1)$. Impose $\mathscr{C}^1$ contact at θ_1, and at some $\theta_2 > \theta_1$ solve

$$h(\theta_1) = f(\theta_1), \quad h'(\theta_1) = f'(\theta_1), \quad h(\theta_2) = W(\theta_2), \quad h'(\theta_2) = W'(\theta_2).$$

with 4 unknowns $(a_0, a_1, a_2, \theta_2)$. With the re-parametrisation $t = \theta_2 - \theta_1 > 0$, this reduces to a single root in t that can be found numerically:

$$G_1(t) := f'(\theta_1) \cdot t + \frac{W'(\theta_2) - f'(\theta_1)}{\exp t - 1}(\exp t - 1 - t) - (W(\theta_2) - f(\theta_1)) = 0, \quad t > 0. \tag{3.1}$$

$$G_2(t) := f'(\theta_1) \cdot t + \frac{W'(\theta_2) - f'(\theta_1)}{1 - \exp(-t)}(\exp(-t) - 1 + t) - (W(\theta_2) - f(\theta_1)) = 0, \quad t > 0. \tag{3.2}$$

(See Appendices B.1 and B.2 for the derivation and explicit formulæ for θ_2, a_2, a_1, and a_0 as functions of t.)

Like the Taylor-based ExEL1, ExEL2 preserves the desirable near-centre behaviour of EL whilst delivering the Wald tail outside the convex hull, thereby enabling χ^2-calibrated tests for far-out hypotheses.

3.3 Choice of the cut-off level

Now we show that our correction does not invalidate the result of Qin and Lawless (1994) if it is asymptotically vanishing: the splicing needs to occur so rarely that the splice is never used asymptotically.

Let T_n denote the relevant EL Wilks statistic under $\mathscr{H}_0 : \theta_0 = c$. For the just-identified mean, $T_n := \mathrm{NLELR}(\theta_0)$, and for over-identified estimating equations, $T_n = \mathrm{NLELR}(\theta_0) - \mathrm{NLELR}(\hat{\theta}_{\mathrm{EL}})$ (Qin–Lawless Wilks form).

Define the extrapolated statistic

$$T_n^{\mathrm{Ex}} := \begin{cases} T_n, & T_n \le f_n^{\max}, \\ E_n(T_n), & T_n > f_n^{\max}, \end{cases}$$

[5] A monotone cubic Hermite (Fritsch–Carlson) spline also works once suitable knot values and derivatives are set, but is harder to analyse.

where E_n is any deterministic $\mathscr{C}^1$ non-decreasing continuation (such as ExEL1/ExEL2 because they agree with f up to the splice and then continue quadratically or via the exponential bridge).

The following theorem establishes the sufficient condition for consistency of the proposed correction.

Theorem 1 *If* $f_n^{\max} \xrightarrow{n\to\infty} \infty$, *then* $T_n^{Ex} \xrightarrow[n\to\infty]{d} \chi^2_p$. □

PROOF Since $T_n \xrightarrow[n\to\infty]{d} \chi^2_p$, the sequence T_n is tight. Hence, for any $\varepsilon > 0$, there exists M with $\sup_n \Pr(T_n > M) < \varepsilon$. For large n, $f_n^{\max} \geq M$, so

$$\Pr(T_n^{\mathrm{Ex}} \neq T_n) = \Pr(T_n > f_n^{\max}) \leq \Pr(T_n > M) < \varepsilon \to 0.$$

Therefore, $|T_n^{\mathrm{Ex}} - T_n| \xrightarrow{\mathbb{P}} 0$ and $T_n^{\mathrm{Ex}} \xrightarrow[n\to\infty]{d} \chi^2_p$. ■

Remarkably, the rate at which $f_n^{\max}$ must tend to ∞ is immaterial for the first-order Wilks property. Smoothness, convexity, or a particular tail shape are also not required for the extrapolated part to possess the Wilks property. Indeed,

$$\Pr(\mathrm{NLELR}(\theta_0) > f_n^{\max} \text{ or } \mathrm{NLELR}(\hat{\theta}_{\mathrm{EL}}) > f_n^{\max}) = o(1) \text{ as } n \to \infty,$$

which gives the χ^2_p limiting distribution for $T_n = \mathrm{NLELR}(\theta_0) - \mathrm{NLELR}(\hat{\theta}_{\mathrm{EL}})$. In particular, despite the fact that ExEL1 does not correspond to a Wald parabola with χ^2_p tails for a fixed n, we have that $T_n^{\mathrm{ExEL1}} \xrightarrow[n\to\infty]{d} \chi^2_p$.

If $T_n = \mathrm{NLELR}(\hat{\theta}_{\mathrm{EL}})$, then the limiting distribution for T_n^{Ex} is χ^2_{q-p} (Qin & Lawless, 1994, Corollary 4).

In practice, the chosen $f_n^{\max}$ as a percentile of the χ^2_p distribution should drift towards 100% as $n \to \infty$. We propose using

$$f_n^{\max} = Q_{\chi^2_p}(1 - \alpha_n), \quad \alpha_n \to 0.$$

Then $\Pr(T_n > f_n^{\max}) \leq \alpha_n + o(1) \to 0$, so the splice is used with vanishing probability and the χ^2_p limit is preserved.

As for the growth rate of $f_n^{\max}$, there is no minimal rate beyond $f_n^{\max} \to \infty$. Here, α_n can go to 0 arbitrarily slowly. Since the upper tail of a χ^2_p variable is proportional to $x^{p/2-1}\mathrm{e}^{-x/2}/\Gamma(p/2)$, using $\alpha_n = n^{-c}$ for $c > 0$ yields $f_{\max}(n) = Q_{\chi^2_p}(1 - n^{-c}) \approx 2c\log n + (p-2)(\log(c\log n)) - 2\log\Gamma(p/2)$.[6]

For second-order refinements, e. g. Bartlett correction, it is convenient (but not strictly necessary for the first-order Wilks property) to make the mass that is modified go to zero fast enough. A sufficient condition for the vanishing mean distortion is $\alpha_n f_n^{\max} \to 0$ because for a non-negative $T \sim \chi^2_p$, tail moments obey $\mathbb{E}(T^m \mathbb{I}(T > c)) \sim \alpha(c)c^m$ as $c \to \infty$, where $\alpha(c) := \Pr(T > c)$.[7] Thus the mean distortion from splicing only when $T > f_n^{\max}$ is of the order $\alpha_n f_n^{\max}$. To make this quantity negligible relative to the n^{-1} scale of Bartlett corrections, one needs $\alpha_n f_n^{\max} = o(n^{-1}) \Rightarrow \alpha_n n f_n^{\max} \to 0$. To keep second-order effects at the $o(n^{-1})$ level, impose the stronger (but still mild) $\alpha_n (f_n^{\max})^2 = o(n^{-1})$. With $\alpha_n = n^{-c}$ and $f_n^{\max} \sim 2c\log n$, any $c > 1$ satisfies both properties because $n^{-c}(\log n)^2 = o(n^{-1})$.

[6] On a machine with limited precision, expressions of the form $Q_{\chi^2_p}(1 - n^{-c})$ may be evaluated inaccurately due to rounding errors when $n \to \infty$ because $1 - \epsilon/2 \approx 1$, where ϵ is the machine epsilon. In those extreme cases, the approximate upper-quantile formula may be used. When $p = 1$, $Q_{\chi^2_1}(1 - n^{-c}) \approx 2c\log n - \log(c\log n) - \log\pi$, and when $p = 2$, $Q_{\chi^2_2}(1 - n^{-c}) = 2c\log n$.

[7] This is because $\mathbb{E}(T^m \mathbb{I}(T > c)) = 2^m \Gamma(p/2 + m, c/2)/\Gamma(p/2)$, where $\Gamma(\cdot)$ is the gamma function and $\Gamma(\cdot,\cdot)$ is the upper incomplete gamma function. The survival function is $\alpha(c) = \Pr(T > c) = \Gamma(p/2, c/2)/\Gamma(p/2)$. Hence $\mathbb{E}(T^m \mathbb{I}(T > c))/(c^m \alpha(c)) = \Gamma(p/2 + m, c/2)/((c/2)^m \Gamma(p/2, c/2)) := R_m(c)$. A recursive argument shows that $R_{m+1}(c) = \frac{p/2+m}{c/2} R_m(c) + (c/2)^{p/2-1}\mathrm{e}^{-c/2}/\Gamma(p/2, c/2)$. As $c \to \infty$, $\Gamma(p/2, c/2) \sim (c/2)^{p/2-1}\mathrm{e}^{-c/2}$, so the second term tends to 1 and the first term tends to 0. With $R_0(c) = 1$, induction gives $R_m(c) \to 1$ for every fixed m. Therefore, $\mathbb{E}(T^m \mathbb{I}(T > c)) \propto \alpha(c)c^m$ as $c \to \infty$, i. e. the m^{th} tail moment is asymptotically c^m times the tail probability.

A practical recommendation follows: choose $\alpha_n = n^{-1-\delta}$ for any $\delta > 0$ (e. g. $n^{-1.1}$) and use the $(1-\alpha_n)$ chi-square quantile as the cut-off level for the splice. This preserves the first-order Wilks property and keeps the splice-induced coverage error below the Bartlett $O(n^{-1})$ term. Inside the hull, the usual Bartlett cut-off inflation still applies.

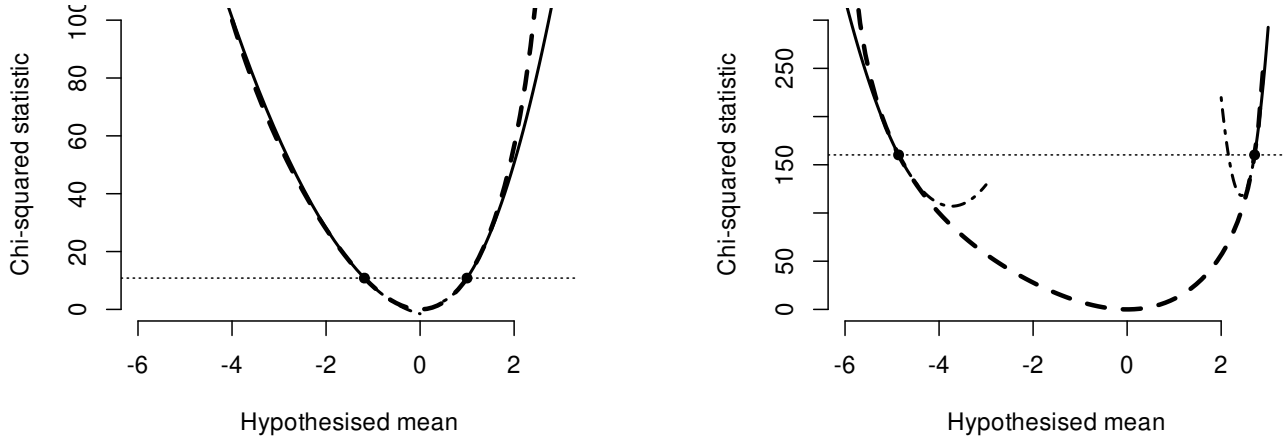
Corollary 1 *If* $T_n = \mathrm{NLELR}(\theta_0) - \mathrm{NLELR}(\hat{\theta}_{EL})$ *is replaced with* $\mathrm{NLELR}(\theta_0) - \mathrm{NLELR}(\hat{\theta}_{ExEL})$*, the Wilks property still holds because* $\hat{\theta}_{EL} = \hat{\theta}_{ExEL}$*.* □

PROOF Since $\hat{\theta}_{\mathrm{EL}}$ minimises $\mathrm{NLELR}(\theta)$, the global minimum does not change: extrapolation is only applied above the threshold $f_n^{\max}$. Therefore, $\hat{\theta}_{\mathrm{EL}}$ and $\hat{\theta}_{\mathrm{ExEL}}$ coincide. ■

Intuitively, this result is due to the fact that splicing takes place only in the upper regions of NLELR, i. e. where it explodes and its first derivative is far from zero. Hence, extrapolation does not affect the regions where the FOC for $\hat{\theta}_{\mathrm{EL}}$ hold.

3.4 Computational aspects and higher-order accuracy

Figure 1 shows two plots demonstrating the behaviour of ExEL for reasonable (left panel, $Q_{\chi^2_1}(1-10^{-3})$) and extreme (right panel, $Q_{\chi^2_1}(1-10^{-36})$) cut-off levels, respectively. When the cut-off level is low, ExEL1 behaves almost like the Wald parabola given by the second-order behaviour of NLELR around zero (see Appendix A.2 for the local expansion). At high cut-offs, NLELR grows super-quadratically near the sample extremes, so the Taylor splice is no longer centred at zero; instead it blends into NLELR at $\theta_{\max}$ and then continues quadratically, while NLELR keeps exploding. This highlights the strengths of ExEL1: it is close to both NLELR and the Wald parabola when the splice is near the centre, yet provides rapidly growing finite values when the splice is near or beyond the sample edge, yielding quasi-statistics in favour of strong rejection of the hypothesis being tested.

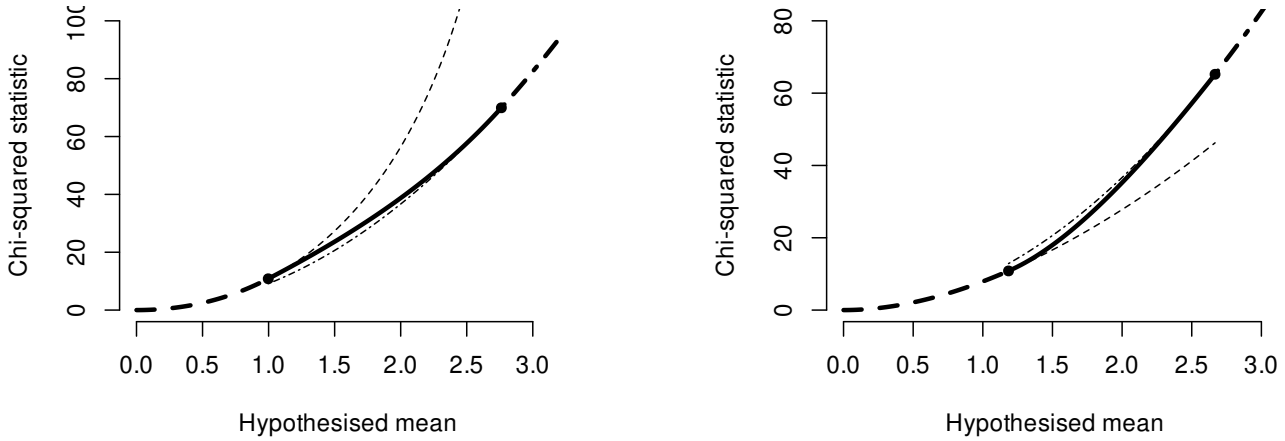


ExEL1 for the uni-variate mean of the sample $Z = -4,-3,\ldots,5$ with weights $10,\ldots,1$. The data are centred around the weighted mean. Extrapolation is used outside the interval $[\theta_{\min},\theta_{\max}]$ where $\theta_{\min},\theta_{\max}$: $\mathrm{NLELR}(\theta_{\min}) = \mathrm{NLELR}(\theta_{\max}) = Q_{\chi^2_1}(0.999) = 10.83$ (left) and $Q_{\chi^2_1}(1-10^{-36}) = 160.24$ (right). Solid: extrapolation; dashed: NLELR; dot-dash: unused branches of the extrapolation parabolas.

Figure 1: Extrapolated EL in one dimension

Figure 2 shows two three-part plots of ExEL2: the original NLELR, the bridge on $[\theta_1,\theta_2]$, and the quadratic tail. In the left panel, NLELR grows faster than the Wald parabola, calling for the h_1 bridge (near-linear at θ_1, exponential by θ_2). In the right panel, the Wald parabola temporarily dominates, calling for the h_2 bridge (U-shaped contact at θ_1, near-linear transition at θ_2).

In terms of algorithmic complexity, ExEL1 can be as fast as the original EL, and ExEL2 has to be slower, but not dramatically so. Recall that the majority of the computational burden in computing $\mathrm{NLELR}(\theta)$ is in finding $\hat{\lambda}$ that yield $\hat{p}_i$. Specifically, it is the damped Newton algorithm (Owen, 2017)



ExEL2 with a bridge $h(\theta)$ to lower (left) and upper (right) parabola at θ_1 computed for the uni-variate mean of the sample $Z = -4, -3, \ldots, 5$ with weights $1, \ldots, 10$ (left) and $10, \ldots, 1$ (right). The data are centred around the weighted mean. Extrapolation is used within the interval $[\theta_1, \theta_2]$ where θ_1: $\mathrm{NLELR}(\theta) = Q_{\chi^2_1}(0.999) = 10.83$ and θ_2 is defined in the text. Solid: bridge; dashed: NLELR; dot-dash: Wald parabola.

Figure 2: Extrapolated EL in one dimension

that uses most of the CPU power: at each step, it evaluates the dual log-likelihood and its two derivatives analytically, computes the search direction, checks the Armijo condition, and backtracks if necessary. In most well-defined EL problems without convex-hull violations, λ iteration counts higher than 10 are practically rare, even in higher dimensions ($q = 20 \ldots 100$).

Lemma 1 *The arithmetic cost of optimising the dual function $\ell_\theta(\lambda) := -\sum_i \log(1 + \lambda' g(Z_i, \theta))$ to find $\hat{\lambda}$ in* (2.4) *is approximately $K \cdot (nq^2 + q^3) + E \cdot nq$ operations, where K is the number of Newton steps bounded via self-concordance and E is the number of total extra function evaluations from backtracking.* □

PROOF In the intended use case, $q \ll n$. The complexity of the damped Newton search is $O(nq^2 + q^3)$ per step, where forming the gradient costs $O(nq)$ evaluations, the Hessian costs $O(nq^2)$ evaluations, and solving a $(q \times q)$ linear system takes $O(q^3)$ operations. The line search costs $O(nq)$ and typically takes only a few trials (typically $E = O(K)$). Self-concordance guarantees global convergence of this backtracking Newton scheme and gives an iteration bound that depends on the initial sub-optimality $-(\ell_\theta(\lambda_0) - \ell_\theta(\hat{\lambda}))$ and the local quadratic convergence near the solution. In practice, one stops when the Newton decrement is less than ε, which corresponds to $-(\ell_\theta(\lambda_K) - \ell_\theta(\hat{\lambda})) \leq \varepsilon^2$. Finally, K is bounded (conservatively) by $O(\log(1/\varepsilon))$, which is a constant function of the user-chosen ε. ■

Once $\hat{\lambda}$ has been found numerically, the derivatives of NLELR that depend on $\hat{\lambda}$ can be computed using two scalars S_0, S_1, a vector T_1, and a matrix T_2 defined in Section 4. These are arithmetic operations that do not increase the algorithmic complexity. Therefore, if $\{\theta_{\min}, \theta_{\max}\}$ are known or were chosen by the user (e. g. based on the 'step away from the boundary by an asymptotically vanishing quantity' 1D rule), then, ExEL1 is as fast as EL. On the other hand, if $\{\theta_{\min}, \theta_{\max}\}$ are to be determined, then, a uni-variate numerical root search may take several steps, each requiring a $\hat{\lambda}$ search, resulting in a small constant multiplicative complexity increase. However, these computational costs may be offset by the ease with which the extrapolated values can be obtained along the same direction.

Finally, ExEL2 introduces one more numerical search for the point at which the values and the slopes match (equation (3.1) or (3.2) for $\mathscr{C}^1$ continuity). At each step of this search, $\hat{\lambda}$ is to be determined, too. This extra search should take no more time than the one for $\theta_{\max}$, making ExEL2 only twice as costly as ExEL1.

In rare cases, a_2 may be positive due to an unusual configuration of NLELR relative to the Wald parabola (e. g. when the hypothesised value is very close to the hull); shifting θ_1 towards the origin remedies this.

Under extreme small-sample configurations (e. g. $n = 5$ with severe outliers), G_1 or G_2 may have multiple roots, potentially breaking convexity. Using the *largest* root of G_1 or G_2 is the desired solution, but guaranteeing it numerically is not trivial. One practical remedy is to move θ_1 inwards until the bridge becomes convex. If numerical instability persists, it is safe to return $W(\theta)$ directly on that side: the conceptual gain from a smooth bridge with unreliable data is negligible.

In small n, convex-hull failure can occur with non-negligible probability, making the ordinary NLELR undefined for many bootstrap resamples. ExEL supplies a finite outer tail via the Taylor or Wald splice, so the bootstrap distribution of NLELR exists for *every* resample (if the latter has at least $q+1$ unique values). Practically, one can run the usual non-parametric bootstrap and compute $-2\log\mathrm{ExELR}$ for each resample; percentile (or studentised) cut-offs then deliver calibrated regions even when the convex-hull condition is violated frequently (where plain EL would drop those resamples or return ∞). This mirrors the motivation behind other 'domain-fixing' approaches like EEL, while ExEL keeps the interior statistic identical to EL; see (Owen, 2001, Section 3.3) for bootstrap calibration of EL.

For the Bartlett correction, let $-2\ell(\theta)$ denote the (interior) EL log-likelihood ratio. For EL there exists a constant b such that $-2(1-b/n)\ell(\theta)$ is χ^2_p up to $O(n^{-2})$ for $\theta = \theta_0$; equivalently, the cut-off can be inflated using $c = Q_{\chi^2_p}(1-\alpha)\cdot(1+b/n)$ (DiCiccio et al., 1991). With ExEL, the simplest choice is threshold calibration: keep the ExEL statistic exactly as constructed (preserving tangency and monotonicity across the splice) and apply the Bartlett-corrected cut-off globally. Inside the hull this reproduces the standard EL Bartlett correction; outside, the Wald tail of ExEL2 already has χ^2_p first-order calibration, and using the same c keeps the region continuous at the splice. When b is unknown, one can estimate it from moments or via a short parametric bootstrap of the EL score; the same estimate can be used to calibrate ExEL cut-offs.

It does not matter that the ExEL1 tails do not correspond directly to any particular χ^2 distribution – asymptotically these tails do not matter for the Bartlett correctability of ExEL1, as we showed in the previous section. The condition that is imposed on the cut-off value ensures that the region on which extrapolation occurs shrinks sufficiently quickly.

4 Multi-variate extension via radial reduction

In multiple dimensions, the problem remains the same: a meaningful extrapolation of ELR beyond the multi-variate convex hull of the data. In this section, we extend the uni-variate approach from Sections 3.1 and 3.2 to handle cases where $\mathrm{NLELR}(\theta)$ becomes a function of a vector θ.

Although for uni-variate EL the convex hull of the data is trivially the interval between two points (minimum and maximum order statistics), for multi-variate EL, the problem becomes harder because the convex hull is not easy to infer from the data; the reasons for avoiding the search for the observations on the hull boundary are given in footnote 4. This is why we reduce a multi-variate problem to a uni-variate one.

Let $Z_1,\dots,Z_n \in \mathbb{R}^p$ be an IID sample with average $\bar{Z}_n$, and suppose that a researcher wishes to test $\mathscr{H}_0\colon \mathbb{E}Z = \theta \in \mathbb{R}^p$ using empirical likelihood for the mean. (The case of testing $\mathbb{E}g(Z,\theta_0) = 0$ is handled similarly.)

The first step is centring the data by subtracting $\bar{Z}_n$ for a clearer exposition; assume without loss of generality that $\sum_{i=1}^n Z_i = 0$. Use the previously defined negative log-empirical likelihood ratio NLELR, denoting it by f for brevity:

$$f(\theta) := -2\log\mathrm{ELR}(\theta) = 2\sum_{i=1}^{n}\log u_i(\theta),$$

where $u_i(\theta) := 1 + \lambda(\theta)'(Z_i - \theta)$ and $\lambda(\theta) \in \mathbb{R}^p$ is the unique Lagrange multiplier solving the EL constraint $\sum_{i=1}^n u_i(\theta)^{-1}(Z_i - \theta) = 0$. On the interior of the convex hull of $Z_1, \ldots, Z_n$, the map f is $\mathscr{C}^2$ and strictly convex (the proof is given in Appendix C.3); $f(\theta) \to \infty$ as θ approaches the hull boundary.

The central idea for higher dimensions is *radial reduction*: along any fixed direction $v \in \mathbb{S}^{p-1}$ the restriction of f to the ray $t \mapsto \theta_v(t) = tv$ behaves as a one-dimensional strictly convex function with the same structure as in Sections 3.1–3.2: EL-like near the centre, explosively growing at the boundary. This lets one reuse ExEL1 and ExEL2 verbatim, provided that they compute derivatives and (for ExEL2) the appropriate directional Wald parabola.

Fix $v \in \mathbb{S}^{p-1}$ and consider the ray

$$\theta_v(t) = tv, \quad t \in (t_{\min}(v), t_{\max}(v)),$$

where $(t_{\min}(v), t_{\max}(v))$ is the maximal open interval for which $\theta_v(t)$ remains in the interior of the hull. Define the uni-variate directional function $f_v(t) := f(\theta_v(t))$. Strict convexity of f implies that f_v is strictly convex on $(t_{\min}(v), t_{\max}(v))$, with $f_v(t) \to +\infty$ as $t \to t_{\max}(v)$. Near $t = 0$, we have the quadratic expansion

$$f_v(t) = n(v'\hat{\Sigma}^{-1}v)\,t^2 + O(t^3),$$

which follows from $f(\theta) = n\theta'\hat{\Sigma}^{-1}\theta + O(\|\theta\|^3)$ at $\theta \approx 0$, with $\hat{\Sigma} := \frac{1}{n}\sum_{i=1}^n Z_i Z_i'$ the sample covariance and $\mathrm{Var}(v'\bar{Z}_n) = v'\hat{\Sigma}v/n$. (See Appendix C.1 for a derivation and Appendix C.2 for the directional expansion.)

For numerical work, it is convenient to have directional derivative formulæ that avoid differentiating λ explicitly. Write (in shorthand notation; full derivations in Appendix C.4)

$$S_0 := \sum_i \frac{1}{u_i(\theta_v)}, \quad S_1 := \sum_i \frac{1}{u_i(\theta_v)^2}, \quad T_1 := \sum_i \frac{Z_i - \theta_v}{u_i(\theta_v)^2}, \quad T_2 := \sum_i \frac{(Z_i - \theta_v)(Z_i - \theta_v)'}{u_i(\theta_v)^2}.$$

Then, along any ray v, the directional derivative of f and of the Lagrange multiplier are

$$f_v'(t) = -2S_0\lambda' v \quad \text{and} \quad \dot{\lambda} := \frac{\mathrm{d}\lambda}{\mathrm{d}t} = T_2^{-1}\big[-S_0 v + (\lambda' v)T_1\big]. \tag{4.1}$$

From these, one obtains a compact form for the second directional derivative:

$$f_v''(t) = 2(w_v' T_2^{-1} w_v) - 2(\lambda' v)^2 S_1,$$

where $w_v(t) := S_0 v - (\lambda' v)T_1$. Then, $\dot{\lambda} = -T_2^{-1}w_v$. By Cauchy–Schwarz in the weighted inner product, the scalar $S_1 - T_1' T_2^{-1} T_1$ is non-negative. Together with $T_2 > 0$ on the interior, this and the full Hessian representation imply that $f_v''(t) > 0$, unless the sample is degenerate.

In practice, these analytic formulæ for f_v' and f_v'' along any ray should be preferred to the fragile finite differences that may be ill-conditioned near the hull.

For ExEL2, we also need the Wald quadratic in direction v. Since $\hat{\Sigma} := \frac{1}{n}\sum_{i=1}^n Z_i Z_i'$ for centred data, the Wald statistic for p-dimensional means is

$$W(\theta) = n\theta'\hat{\Sigma}^{-1}\theta.$$

Restricting to $\theta_v(t) = tv$ gives

$$W_v(t) = n(v'\hat{\Sigma}^{-1}v)\,t^2 := a(v)\,t^2.$$

All the uni-variate splice formulæ in Section 3.2 therefore apply directly with a replaced by $a(v)$.

4.1 Multi-variate algorithm adaptation

Fix the target θ. Write $v = \frac{\theta - \bar{Z}_n}{\|\theta - \bar{Z}_n\|}$ and $t_\theta = \|\theta - \bar{Z}_n\|$. Work with $f_v(t) = f(\bar{Z}_n + tv)$. With centred data, the formulæ simplify to $t_\theta = \|\theta\|$, $v = \theta / t_\theta$, and $f_v(t) = f(tv)$.

ExEL1 becomes the following algorithm.

1. Choose a cut-off level $f_n^{\max}$ (e. g. the $(1 - 1/n^{1+\delta})$ quantile of χ_p^2 for some $\delta > 0$) and find the unique $t_{\max} > 0$ with $f_v(t_{\max}) = f_n^{\max}$. Uni-modality of $f_v(t)$ guarantees uniqueness of $t_{\max}$. In practice, bracket by monotone expansion from 0; a single bisection after first crossing typically suffices.
2. Evaluate f_v, f_v', f_v'' at $t_{\max}$ using the analytic formulæ above (or finite differences if f is a black box and λ is unavailable).
3. Form the directional Taylor polynomial

$$T_v(t) = f_v(t_{\max}) + f_v'(t - t_{\max}) + \frac{1}{2} f_v''(t - t_{\max})^2$$

   and use it as the extrapolated branch for $t > t_{\max}$.

The resulting spliced function is $\mathscr{C}^1$ at the splice radius in each direction.

ExEL2 becomes the directional Wald splice.

Construct a strictly convex bridge by choosing any $t_1 > 0$ (e. g. the same cut as in ExEL1). Set $t_2 = t_1 + \tau$, and check if $f_v(t_1) > W_v(t_1)$; if yes, solve G_{v1} for $\tau > 0$, otherwise solve G_{v2} for $\tau > 0$:

$$G_{v1}(\tau) := f_v'(t_1)\tau + \frac{W_v'(t_2) - f_v'(t_1)}{\exp\tau - 1}(\exp\tau - \tau - 1) - (W_v(t_2) - f_v(t_1)) = 0,$$

$$G_{v2}(\tau) := f_v'(t_1)\tau + \frac{W_v'(t_2) - f_v'(t_1)}{1 - \exp(-\tau)}(\exp(-\tau) + \tau - 1) - (W_v(t_2) - f_v(t_1)) = 0,$$

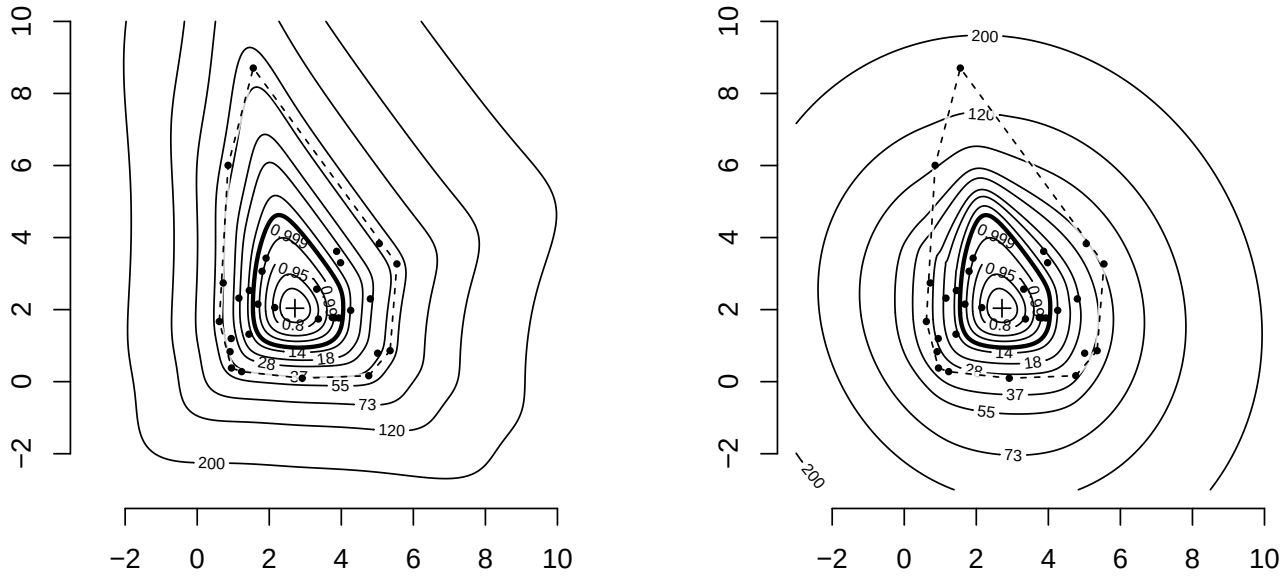
which yields a $\mathscr{C}^1$ splice. Enforce $\tau \geq \max(0, f_v'(t_1)/(2a(v)) - t_1)$ so that the bridge curvature coefficient be positive (hence strict convexity). Existence is guaranteed by $G_v(0^+) > 0$ and $G_v(t) \sim -a(v)t^2$ as $t \to \infty$ for both G_{v1} and G_{v2}.

4.2 Regularity of the multi-variate extrapolation

Let $C := \{\theta\colon f(\theta) \leq c\}$ denote EL sub-level sets; C are convex for the mean EL. Along each ray v, the equation $f_v(t) = c$ has at most one solution on the positive branch, so each ray intersects the boundary of C once. As v varies, the cut radius $t_{\max}(v)$ and the ExEL2 splice radii vary *continuously* under mild regularity (continuity of f_v and the implicit function theorem away from grazing directions). Hence the assembled extrapolated surface $\theta \mapsto f_{**}(\theta)$ is continuous and piecewise $\mathscr{C}^1$ across angles; it is fully $\mathscr{C}^1$ wherever the regime (line bridge vs. exponential bridge) does not change with v. In practice, transitions between regimes occur on a thin angular set determined by $f_v''(0) = 2a(v)$.

ExEL2 yields convex sets – ellipses – for extreme hypothesised values of θ. For ExEL1, radial convexity along rays does not imply global convexity: the varying curvature and derivatives introduce potential non-convex sub-level sets. Convexity of all sub-level sets would require the function to be convex in θ, which radial Taylor pieces do not ensure when their coefficients vary with direction v. Hence, using a Taylor expansion might not preserve convexity across directions, resulting in inward dents of level sets for extreme levels.

Figure 3 shows the contours produced by ExEL1 and ExEL2. Since the data used for the plot are χ_3^2-distributed, the general shape of the convex hull resembles the contour level of the underlying density. ExEL1 is not only faster but also better at preserving the shape of the contour levels for extreme hypotheses not supported by the data. ExEL2 requires more function evaluations because



ExEL1 (left) and ExEL2 (right) computed for the bi-variate mean of a random χ^2_3 sample with 30 observations and random uniform [0, 1] weights. Extrapolation is used outside the region θ: $\mathrm{NLELR}(\theta) \leq Q_{\chi^2_2}(0.999) = 13.82$. Dashed line: convex hull of the data. Thick line: boundary of the level set containing original EL. The cross is the weighted average.

Figure 3: Extrapolated EL in two dimensions

it has to determine both θ_1 and θ_2 for splicing, which results in two numerical searches. Its resulting shapes are also similar to those of the set corresponding to the chosen level $Q_{\chi^2_2}(0.999) = 13.82$, but as the level increases, the shape is morphing into an ellipse. Remarkably, the contour levels in the transition zone can be marginally non-convex even for ExEL2 despite the fact that a convex bridge is constructed along each ray between two points at which the inner and outer level sets are convex.

Extrapolated EL is not without its pitfalls; certain measures must be taken to ensure that its performance is satisfactory.

1. If the cloud $Z_1, \ldots, Z_n$ lies (nearly) in a lower-dimensional affine subspace, $T_v(\theta)$ may be ill-conditioned in some directions since it depends on T_2; Newton solver for $\lambda(\theta)$ and for the exponential-bridge should then be warm-started.
2. All evaluations of $f_v(t)$ and its derivatives must keep $u_i(\theta_v(t)) > 0$. Bracketing and step-size control along the ray help avoid accidental excursions.
3. For a two-dimensional problem, an initial caching of parabola coefficients (for ExEL1) or bridge coefficients (for ExEL2) on a coarse angular grid with interpolation over v for an approximate surface allows near-instant evaluation during optimisation and plotting.

5 Simulations

5.1 Mean coverage probability

We evaluate the small-sample accuracy of both ExEL variants against the ordinary empirical likelihood (EL), adjusted EL (AEL, J. Chen et al., 2008), and balanced EL (BEL, Emerson and Owen, 2009) in the classical mean-estimation problem. Following DiCiccio et al. (1991) and subsequent EL-modification studies (e. g. Baragona et al., 2017; Bartolucci, 2007; Y. Liu and Chen, 2010; Tsao, 2013; Tsao and Wu, 2013, 2014), we consider settings where the convex-hull restriction is most likely to be an obstacle that needs overcoming.

We assess finite-sample coverage probability for the mean under the classical design (J. Chen et al., 2008; DiCiccio et al., 1991). In all experiments the null hypothesis is that the true population

mean is $\mathbb{E}Z = \theta_0 = 0$. The aforementioned papers study three distributional families: Gaussian $\mathcal{N}(0,1)$ with $n \in \{10,20\}$, shifted chi-squared $\chi^2_1 - 1$ with $n \in \{20,40\}$, and Student-t with $n \in \{15,30\}$. We extend the simulation by considering extremely small samples (the smallest sample size from above halved and rounded down) and a heavy-tailed distribution with all finite moments – centred log-normal, $\log Z - \exp(1/2)$ where $\log Z \sim \mathcal{N}(0,1)$, with $n \in \{5,10,20\}$.

For each setting, we run 10 000 Monte-Carlo replications. In replication m, a sample $\{Z_i^{(m)}\}_{i=1}^n$ is generated and the following statistics are computed at θ_0:

- EL: the original negative log-empirical likelihood ratio (NLELR) $-2\log\mathrm{ELR}(\theta_0)$; when the spanning condition fails, the implementation returns $\log\mathrm{ELR}(\theta_0) = -\infty$, and we record the chi-square statistic as $+\infty$ (a rejection of $\mathcal{H}_0$ for that replication);
- ExEL1 / ExEL2: the proposed extrapolated NLELRs. In the code we compute both on every sample. By design, ExEL is finite for all θ, so the test can make a meaningful numerical comparison even when the original EL fails due to the convex-hull condition. In all simulations, we used a root solver to find the cut-offs (denoted $\theta_{\min}, \theta_{\max}$ for ExEL1 and $(\theta_{1L}, \theta_{1R})$ for ExEL2) such that the NLELR value at the cut-off – i. e. $f_n^{\max}$ – equals $Q_{\chi^2_p}(1 - n^{-1.1})$;
- AEL / BEL: adjusted and balanced EL, respectively. In 1 and 2 dimensions, we use the coefficients on the pseudo-observations due to Y. Liu and Chen (2010) as they seem to provide better accuracy than the original proposal due to J. Chen et al. (2008).

We test at four nominal levels $(1-\alpha) \in \{0.80, 0.90, 0.95, 0.99\}$ using three calibrations:

- Plain χ^2: reject when the statistic exceeds $Q_{\chi^2_p}(1-\alpha)$.
- Bartlett adjustment: we estimate the Bartlett factor in two ways: (1) using the bias-adjusted moment estimator of Y. Liu and Chen (2010) and (2) via bootstrap. We set $\hat{c} = 1 + \hat{b}/n$ and compare $\hat{c}^{-1}\,\mathrm{NLELR}(\theta_0)$ to $Q_{\chi^2_p}(1-\alpha)$. We use a robust modification of the bootstrap Bartlett-factor estimator described in Appendix E.
- Bootstrap calibration (percentile of the statistic): for each data set, we compute the bootstrap distribution of the finite (owing to the ExEL1 extrapolation) statistic $\mathrm{NLELR}^*(\bar{Z}_n)$ with $B = 1000$ resamples and compare the observed statistic to the empirical quantiles. If only one unique observation remains (or two in 2D) and a bootstrap resample is degenerate, we redraw.

For each replication we also record an indicator of EL failure (convex-hull violation). The convex-hull failure rate we report is simply the share of replications in which ELR is not finite; in such cases ExEL returns a finite statistic (so the replication is 'redeemed'). Our second operational measure is the share of bootstrap resamples in which the convex-hull condition fails.

For each 'distribution – sample-size' pair we report:

- Empirical rejection probabilities for the original EL (where defined), ExEL1, ExEL2, AEL, and BEL under plain χ^2 calibration (first three rows);
- The same after moment-based ('momBart') and bootstrap-based Bartlett correction of two kinds: redraw until the convex-hull condition holds ('$\mathrm{ELCH}_{\mathrm{bootBart}}$'), and ExEL1-based bootstrap Bartlett ('$\mathrm{ExEL1}_{\mathrm{bootBart}}$');
- The same after bootstrap calibration of two kinds: (1) redraw until the convex-hull condition holds ('$\mathrm{ELCH}_{\mathrm{boot}}$'), and (2) ExEL1-based ('$\mathrm{ExEL1}_{\mathrm{boot}}$').

We also report three rates: (1) fraction of EL failures in the full sample, (2) fraction of EL failures in bootstrap samples, and (3) fraction of bootstrap resamples in which extrapolation was used (regardless of EL failure).

The key takeaways from Tables 1 and 2 are as follows.

Under plain χ^2_1 calibration, ExEL1/ExEL2 and EL have *essentially the same coverage* in 1D. This is expected: inside the hull, ExEL equals EL; when the hull fails, ExEL gives a large finite statistic, but the comparison is still made to the same χ^2_1 cut-off, so rejection behaviour does not change

Panel 1: Normal distribution

	$n=5$				$n=10$				$n=20$			
Nominal	0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99
EL	.664	.752	.809	.869	.750	.846	.901	.957	.773	.876	.932	.980
ExEL1	.664	.752	.811	.881	.750	.846	.901	.958	.773	.876	.932	.980
ExEL2	.664	.755	.819	.902	.750	.846	.901	.960	.773	.876	.932	.980
AEL	.702	.786	.842	.902	.790	.884	.933	.977	.794	.892	.943	.986
BEL	.800	.905	.954	.991	.810	.910	.957	.993	.798	.902	.951	.990
$\text{ExEL1}_{\text{momBart}}$	.702	.788	.839	.901	.781	.873	.921	.968	.792	.890	.941	.984
$\text{ELCH}_{\text{bootBart}}$	.697	.777	.826	.878	.789	.879	.925	.968	.788	.888	.939	.984
$\text{ExEL1}_{\text{bootBart}}$	.783	.853	.892	.937	.792	.883	.928	.971	.788	.888	.939	.984
$\text{ELCH}_{\text{boot}}$	.707	.782	.845	.890	.791	.891	.941	.984	.788	.893	.945	.989
$\text{ExEL1}_{\text{boot}}$	.787	.889	.977	.997	.796	.897	.948	.989	.788	.893	.945	.989
Rates, %	6.5 / 10.4 / 29.6				0.2 / 0.6 / 12.9				0.0 / 0.0 / 5.1			

Panel 2: χ^2_1 distribution

	$n=10$				$n=20$				$n=40$			
Nominal	0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99
EL	.677	.777	.836	.900	.734	.837	.893	.955	.755	.860	.915	.971
ExEL1	.677	.777	.836	.901	.734	.837	.893	.955	.755	.860	.915	.971
ExEL2	.677	.777	.834	.893	.734	.837	.893	.952	.755	.860	.915	.970
AEL	.775	.855	.895	.943	.789	.876	.922	.967	.783	.877	.930	.976
BEL	.747	.831	.878	.935	.751	.857	.911	.965	.764	.865	.922	.975
$\text{ExEL1}_{\text{momBart}}$	.739	.824	.870	.921	.773	.864	.914	.963	.778	.873	.927	.976
$\text{ELCH}_{\text{bootBart}}$	.728	.817	.866	.917	.769	.863	.912	.963	.778	.873	.927	.976
$\text{ExEL1}_{\text{bootBart}}$	.753	.836	.878	.926	.771	.864	.912	.963	.778	.873	.927	.976
$\text{ELCH}_{\text{boot}}$	.734	.834	.891	.939	.773	.875	.928	.976	.779	.879	.935	.980
$\text{ExEL1}_{\text{boot}}$	.752	.864	.911	.966	.774	.876	.929	.976	.779	.879	.935	.981
Rates, %	2.1 / 3.3 / 17.5				0.0 / 0.2 / 8.3				0.0 / 0.0 / 3.6			

Rates: (1) CH failure rate in the full sample, (2) CH failure rate in bootstrap samples, (3) extrapolation rate in bootstrap samples (regardless of the CH failure).

Table 1: Coverage probabilities of population mean in one dimension for light-tailed distributions

Panel 3: t_5 distribution

	$n=7$				$n=15$				$n=30$			
Nominal	0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99
EL	.691	.794	.856	.926	.744	.853	.911	.968	.776	.881	.935	.983
ExEL1	.691	.794	.856	.930	.744	.853	.911	.968	.776	.881	.935	.983
ExEL2	.691	.795	.862	.940	.744	.853	.911	.970	.776	.881	.935	.983
AEL	.781	.873	.921	.971	.792	.891	.941	.984	.799	.897	.948	.988
BEL	.810	.919	.965	.995	.794	.902	.953	.992	.790	.895	.947	.990
$\text{ExEL1}_{\text{momBart}}$	.743	.838	.892	.949	.777	.877	.930	.978	.795	.894	.946	.986
$\text{ELCH}_{\text{bootBart}}$	.751	.840	.891	.944	.782	.880	.931	.978	.794	.892	.943	.986
$\text{ExEL1}_{\text{bootBart}}$	.789	.872	.916	.961	.782	.881	.932	.979	.794	.892	.943	.986
$\text{ELCH}_{\text{boot}}$	.758	.859	.914	.957	.784	.892	.945	.988	.795	.896	.949	.989
$\text{ExEL1}_{\text{boot}}$	.791	.903	.956	.996	.784	.893	.946	.989	.795	.896	.949	.989
Rates, %	1.5 / 4.0 / 22.0				0.0 / 0.1 / 8.8				0.0 / 0.0 / 3.6			

Panel 4: Log-normal(0, 1) distribution

	$n=5$				$n=10$				$n=20$			
Nominal	0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99
EL	.562	.647	.700	.763	.641	.745	.805	.878	.697	.804	.870	.935
ExEL1	.562	.647	.703	.775	.641	.745	.805	.880	.697	.804	.870	.934
ExEL2	.562	.650	.709	.780	.641	.745	.804	.873	.697	.804	.870	.929
AEL	.635	.711	.759	.822	.739	.819	.865	.920	.771	.857	.901	.949
BEL	.688	.771	.831	.917	.717	.806	.858	.926	.717	.827	.887	.947
$\text{ExEL1}_{\text{momBart}}$	.615	.692	.741	.801	.711	.792	.842	.900	.751	.844	.893	.946
$\text{ELCH}_{\text{bootBart}}$	.562	.647	.703	.765	.697	.784	.836	.897	.747	.841	.891	.944
$\text{ExEL1}_{\text{bootBart}}$	.699	.763	.798	.846	.726	.806	.851	.906	.751	.843	.892	.944
$\text{ELCH}_{\text{boot}}$	.580	.657	.728	.782	.703	.804	.865	.926	.751	.857	.908	.960
$\text{ExEL1}_{\text{boot}}$	.705	.791	.916	.974	.728	.831	.884	.950	.754	.859	.909	.960
Rates, %	16.0 / 14.5 / 31.2				2.2 / 3.6 / 18.1				0.0 / 0.3 / 9.6			

Rates: (1) CH failure rate in the full sample, (2) CH failure rate in bootstrap samples, (3) extrapolation rate in bootstrap samples (regardless of the CH failure).

Table 2: Coverage probabilities of population mean in one dimension for heavy-tailed distributions

systematically. Small discrepancies arise because the extrapolated branch grows quadratically (ExEL) versus explosively (EL), so very early cut-offs (80% here) can shift a few outcomes.

In small samples, EL strongly under-covers (e. g. normal, $n = 5$: 0.95-level coverage is ≈0.81). AEL and BEL provide better coverage; BEL is superior to AEL almost everywhere (except for the χ^2_1 design). Coverage improves with n, as expected. With heavy tails / skew (log-normal), under-coverage of the χ^2 calibration is severe (e. g. $n = 5$: 0.95-coverage is ≈0.70) and convex-hull failure can be non-negligible (16% at $n = 5$). In these cases ExEL is crucial because it defines a finite test statistic for every sample and enables resampling-based calibration.

Moment-based Bartlett correction is computationally lightweight (< 1 ms) and produces near-optimal results for moderate n and light tails (e. g. normal when $n = 20$: 0.95-coverage is ≈0.94), but it is less corrective than the bootstrap in heavy-tailed/small-n settings (log-normal, $n = 5$). Bootstrap Bartlett corrections tend to be insufficient both in tiny, well-behaved samples (e. g. normal, $n = 5$) and under severe skew; by $n = 20$ they move close to both moment-based corrections and percentile bootstrap rows.

Bootstrap calibration (percentile) is the most effective overall in the hard regimes (small n, heavy tails), pulling coverage toward nominal and often closest to target – though in the smallest normal designs it can be slightly conservative.

The danger of discarding failed bootstrap replications is evident when comparing the 'ELCH' rows (redraw until the spanning condition holds) with the ExEL-based bootstrap variants: coverage under ELCH is systematically lower – often markedly so – than under ExEL bootstrap, indicating over-rejection of the null caused by truncating the upper tail of the statistic.

Convex-hull failure rates (bottom lines) are non-negligible exactly where coverage is poor (small n, heavy tails), and near zero elsewhere. ExEL does not inflate size by itself under plain χ^2 cut-offs, but it rescues those failed draws so one can (1) report finite statistics, (2) do bootstrap calibration on every sample, and (3) apply Bartlett uniformly – hence the improved calibrated rows.

Finally, since the chosen $f_n^{\max}$ increases as n grows, the extrapolation rate in bootstrap samples goes down with increasing n.

The picture is almost similar for bi-variate data (Tables 3 and 4). Under plain χ^2 cut-offs, ExEL1/ExEL2 essentially coincide with EL (same rows), as expected: inside the hull they are identical; outside, ExEL returns a large finite value but we compare to the same χ^2 threshold, so raw coverage probability hardly changes. AEL and BEL provide a marginal improvement over the plain χ^2_2 calibration, but it is insufficient for reliable inference. Moment-based Bartlett makes steady, small improvements toward nominal but cannot fully overcome small-sample distortions. Bootstrap Bartlett is again too mild in all settings, but better than moment-based Bartlett, converging toward sensible corrections as n grows. Bootstrap calibration is consistently the best (closest to nominal) in the hard regimes (small n, skew/heavy tails), and remains competitive for moderate n. Convex-hull failure rates are higher in 2D at a given n; this is exactly where ExEL is operationally crucial: it enables bootstrapping on every replication.

Therefore, we advocate percentile bootstrap calibration using ExEL so every resampled data set is admissible when accuracy at small n or under non-Gaussianity matters. Moment-based Bartlett is a good fast default for moderate n and well-behaved distributions.

5.2 Regression parameter estimation under conditional moment restrictions

We study estimation of the parameter vector $\theta_0 := (\alpha_0, \beta_0, \gamma_0, \delta_0)$ in the non-linear regression model

$$Y = \alpha_0 + \beta_0 X_1 + \gamma_0 X_2^{\delta_0} + U, \tag{5.1}$$

where $\mathbb{E}(U \mid X) = 0$, $(X_1, X_2) \overset{\text{IID}}{\sim} \exp(1)$, and the model error is mildly heteroskedastic and heavy-tailed with $U/\sqrt{1 + X_1 + X_2} \sim t_5$. We set $\alpha_0 = 0$, $\beta_0 = \gamma_0 = 1$, $\delta_0 = 0.5$ and $n = 100$ observations.

Panel 5: Normal distribution

	$n = 10$				$n = 20$				$n = 50$			
Nominal	0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99
EL	.460	.611	.712	.841	.505	.686	.799	.927	.540	.720	.835	.954
ExEL1	.460	.611	.712	.841	.505	.686	.799	.927	.540	.720	.835	.954
ExEL2	.460	.611	.712	.843	.505	.686	.799	.927	.540	.720	.835	.954
AEL	.506	.661	.762	.894	.537	.717	.826	.941	.553	.734	.846	.958
BEL	.555	.741	.850	.966	.547	.738	.848	.961	.552	.735	.849	.961
$\text{ExEL1}_{\text{momBart}}$	.497	.651	.747	.863	.535	.714	.822	.938	.552	.734	.845	.958
$\text{ELCH}_{\text{bootBart}}$	.562	.707	.793	.893	.550	.730	.834	.944	.555	.737	.848	.958
$\text{ExEL1}_{\text{bootBart}}$	.613	.752	.831	.919	.550	.731	.835	.944	.555	.737	.848	.958
$\text{ELCH}_{\text{boot}}$	.764	.872	.922	.962	.790	.892	.948	.989	.796	.897	.947	.989
$\text{ExEL1}_{\text{boot}}$	.812	.919	.969	.997	.790	.893	.948	.990	.796	.897	.947	.989
Rates, %	2.0 / 5.0 / 22.7				0.0 / 0.0 / 7.4				0.0 / 0.0 / 1.9			

Panel 6: χ^2_1 distribution

	$n = 10$				$n = 20$				$n = 50$			
Nominal	0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99
EL	.368	.503	.593	.721	.455	.613	.727	.867	.515	.687	.799	.926
ExEL1	.368	.503	.593	.721	.455	.613	.727	.867	.515	.687	.799	.926
ExEL2	.368	.503	.593	.723	.455	.613	.727	.867	.515	.687	.799	.926
AEL	.408	.550	.647	.781	.489	.656	.766	.891	.540	.712	.816	.935
BEL	.458	.612	.705	.833	.482	.653	.770	.901	.530	.699	.811	.932
$\text{ExEL1}_{\text{momBart}}$	.402	.539	.630	.749	.484	.647	.758	.884	.538	.709	.815	.934
$\text{ELCH}_{\text{bootBart}}$	.470	.595	.682	.780	.550	.706	.800	.906	.558	.727	.830	.940
$\text{ExEL1}_{\text{bootBart}}$	.624	.724	.785	.856	.558	.715	.806	.908	.558	.727	.830	.940
$\text{ELCH}_{\text{boot}}$	.657	.756	.810	.861	.767	.876	.928	.977	.785	.886	.938	.986
$\text{ExEL1}_{\text{boot}}$	.763	.886	.937	.985	.773	.882	.934	.983	.785	.886	.938	.986
Rates, %	9.6 / 12.7 / 29.9				0.2 / 0.9 / 13.3				0.0 / 0.0 / 3.7			

Rates: (1) CH failure rate in the full sample, (2) CH failure rate in bootstrap samples, (3) extrapolation rate in bootstrap samples (regardless of the CH failure).

Table 3: Coverage probabilities of population mean in two dimensions for light-tailed distributions

Panel 7: t_5 distribution

	$n = 10$				$n = 20$				$n = 50$			
Nominal	0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99
EL	.427	.583	.686	.825	.488	.662	.778	.916	.523	.705	.823	.949
ExEL1	.427	.583	.686	.825	.488	.662	.778	.916	.523	.705	.823	.949
ExEL2	.427	.583	.686	.830	.488	.662	.778	.916	.523	.705	.823	.949
AEL	.476	.640	.747	.893	.526	.702	.815	.938	.540	.726	.839	.957
BEL	.541	.730	.853	.971	.523	.712	.831	.956	.533	.720	.835	.957
$\text{ExEL1}_{\text{momBart}}$	.468	.625	.724	.851	.521	.695	.806	.930	.540	.725	.837	.957
$\text{ELCH}_{\text{bootBart}}$	.548	.694	.787	.889	.555	.726	.836	.943	.550	.733	.845	.959
$\text{ExEL1}_{\text{bootBart}}$	.617	.754	.834	.920	.556	.727	.837	.943	.550	.733	.845	.959
$\text{ELCH}_{\text{boot}}$	.761	.866	.918	.961	.792	.900	.951	.991	.791	.896	.949	.989
$\text{ExEL1}_{\text{boot}}$	.814	.923	.972	.998	.792	.901	.952	.991	.791	.896	.949	.989
Rates, %	2.0 / 5.8 / 24.9				0.0 / 0.1 / 9.0				0.0 / 0.0 / 2.4			

Panel 8: Log-normal(0, 1) distribution

	$n = 10$				$n = 20$				$n = 50$			
Nominal	0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99
EL	.333	.461	.558	.690	.403	.562	.676	.821	.473	.644	.760	.898
ExEL1	.333	.461	.558	.690	.403	.562	.676	.821	.473	.644	.760	.898
ExEL2	.333	.461	.558	.693	.403	.562	.676	.821	.473	.644	.760	.898
AEL	.374	.513	.613	.755	.447	.605	.717	.851	.501	.671	.783	.915
BEL	.433	.581	.678	.817	.429	.594	.713	.857	.488	.652	.767	.907
$\text{ExEL1}_{\text{momBart}}$	.368	.498	.592	.720	.438	.594	.705	.842	.497	.666	.778	.911
$\text{ELCH}_{\text{bootBart}}$	.440	.569	.650	.757	.529	.675	.767	.875	.544	.710	.815	.927
$\text{ExEL1}_{\text{bootBart}}$	.601	.700	.759	.838	.545	.687	.778	.878	.544	.710	.815	.927
$\text{ELCH}_{\text{boot}}$	.629	.731	.795	.853	.737	.844	.904	.965	.773	.875	.929	.980
$\text{ExEL1}_{\text{boot}}$	.736	.863	.919	.982	.748	.853	.910	.972	.773	.875	.929	.980
Rates, %	10.0 / 12.9 / 30.8				0.4 / 1.3 / 15.7				0.0 / 0.0 / 5.7			

Rates: (1) CH failure rate in the full sample, (2) CH failure rate in bootstrap samples, (3) extrapolation rate in bootstrap samples (regardless of the CH failure).

Table 4: Coverage probabilities of population mean in two dimensions for heavy-tailed distributions

A popular approach for conditional moment restrictions of the form $\mathbb{E}(\rho(Z,\theta_0) \mid X) = 0$ is smoothed empirical likelihood (SEL) (Kitamura et al., 2004). The SEL estimator maximises the sum of local likelihoods with kernel weights w_{ij}:

$$\hat{\theta}_{\text{SEL}} := \arg\max_{\theta} \max_{p_{ij}} \sum_{i,j} w_{ij} \log p_{ij} \quad \text{s.t.} \begin{cases} \sum_{j=1}^{n} p_{ij} = 1 & \forall i = 1,\dots,n \\ \sum_{j=1}^{n} p_{ij}\rho(Z_j,\theta) = 0 & \forall i = 1,\dots,n \end{cases} \tag{5.2}$$

In model (5.1), the residual function $\rho(Z,\theta)$ is $Y - \alpha - \beta X_1 - \gamma X_2^{\delta}$, where $Z := (Y, X_1, X_2)$. Evaluating SEL here amounts to solving n uni-variate locally weighted EL problems; the weights w_{ij} can be interpreted as fractional observation counts (Owen, 2017). In practice w_{ij} come from a kernel with bandwidth b_n; choosing b_n remains an open problem. We adopt the Epanechnikov kernel with an adaptive Loftsgaarden–Quesenberry-style k-nearest-neighbour balloon bandwidth at each centre X_i: for row i, we choose the largest bandwidth $b_n^{(i)}$ such that $w_{ij} > 0$ for exactly 7 indices j (6 neighbours), ensuring each local EL/ExEL problem is well-posed.[8]

A practical difficulty that requires the use of ExEL is the *starting-value problem*: without prior information on effect sizes, it is virtually impossible to manually choose a viable initial value. Starting from all zeros may produce $\rho(Z,0)$ with same-sign values in many neighbourhoods, making the inner EL equal $-\infty$ – which breaks most optimisers. ExEL prevents the $-\infty$ values from appearing owing to finite, convex splices, allowing the global optimiser to run while leaving all feasible neighbourhoods unchanged.

The Monte Carlo design is as follows (10 000 replications):

1. Set the seed and draw (X_1, X_2, U, Y) according to (5.1);
2. Choose the largest $b_n^{(i)}$ such that $\sum_j \mathbb{I}(w_{ij} > 0) = 7$ for each $i = 1,\dots,n$;
3. Maximise (5.2) from five different starting values using BFGS, with the exponent δ constrained to $[0,3]$ and the relative convergence tolerance $1.5 \cdot 10^{-8}$.

We compare five starting values:

1. A vector of zeros: $(0,0,0,0)$;
2. OLS under the mis-specified linear model $\delta_0 = 1$;
3. Smoothed Euclidean likelihood (SEuL) (Antoine et al., 2007, Section 2.2) (also known as Conditional Euclidean Empirical Likelihood – a quadratic approximation to SEL equivalent to the continuously updated SMD of Ai and Chen (2003) with kernel sieves);
4. Non-linear least squares (NLS) (initialised at the OLS estimates);
5. The true parameter $(0,1,1,0.5)$ (oracle; infeasible in practice).

Initialisation	Zero	OLS	SEuL	NLS	Truth
% at least one fail	100.0	51.9	32.0	54.1	49.7
Median iterations	34	29	24	31	32
% diverged	26.7	25.9	32.4	24.8	27.5

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

Table 5: Optimisation diagnostics for SEL

[8] Operationally, $b_n^{(i)}$ is the L_∞ distance from X_i to its $(k+1)^{\text{st}}$ nearest neighbour. Our baseline scales as $k \propto \sqrt{n}$, which satisfies the standard consistency conditions $k \to \infty$ and $k/n \to 0$ for k-NN smoothing (Devroye & Wagner, 1977). This rule reflects three principles: adaptivity (larger bandwidth in sparse regions), $\sqrt{n}$ growth of k, and mild under-smoothing to reduce bias (Calonico et al., 2018).

Table 5 shows that whenever the spanning condition fails at the initial value in at least one neighbourhood, EL-based optimisation would not even start. With ExEL, 100% of the runs from the zero start are rescued. Across other starts, ExEL prevents 32–54% of potential initialisation failures. No ExEL-based SEL run fails outright, underscoring the usefulness of ExEL as a robust guard against convex-hull violations at unlucky starting values. Notably, in 26% of simulations all five starts produce at least one non-finite inner EL; without ExEL, these replications would be discarded, inducing selection bias. This mirrors Owen (2001, Section 3.3): if the fraction of resamples with convex-hull failure is substantial, that itself is diagnostic. An analogous phenomenon appears with 'ELCH' bootstrap variants in Tables 1–4: bootstrapping until the convex-hull condition holds drops upper-tail critical values of $-2\log\mathrm{ELR}$; here, it would drop feasible $\hat{\theta}$ values in 'complicated' data sets.

The SEuL start yields the fewest outer iterations, consistent with SEuL's proximity to SEL. Zeros, OLS, NLS, and even the oracle start require more iterations: the finite-sample objective differs from its population counterpart, so the population maximiser is not necessarily an optimal start for the sample problem. We label a run 'diverged' if $\max_k |\hat{\theta}^{(k)} - \theta_0^{(k)}| > 10$; this occurs in 25–32% of runs, plausibly due to the moderate sample size and the complexity of the objective function due to the non-linearity in parameters. In additional simulations (not reported), increasing $b_n^{(i)}$ to 9, 14, or 19 neighbours reduced initial failures (sometimes below 1%, except for the zero initialisation) but slightly raised the percentage of 'diverged' optimisations, suggesting that divergence relates to the objective's shape (through n and $b_n^{(i)}$) and not to the spanning-condition failure at the start.

Initialisation	Zero	OLS	SEuL	NLS	Truth
$\hat{\alpha}$	−0.15 (0.68)	−0.12 (0.73)	−0.31 (0.78)	−0.10 (0.73)	−0.22 (0.68)
($\alpha_0 = 0$)	[−29.4, 0.39]	[−9.22, 0.44]	[−191, 0.37]	[−7.35, 0.46]	[−56.7, 0.32]
$\hat{\beta}$	1.00 (0.21)	1.00 (0.20)	1.00 (0.20)	1.01 (0.20)	1.00 (0.20)
($\beta_0 = 1$)	[0.78, 1.20]	[0.80, 1.21]	[0.80, 1.21]	[0.80, 1.21]	[0.80, 1.21]
$\hat{\gamma}$	1.13 (0.83)	1.05 (0.91)	1.29 (0.91)	1.03 (0.92)	1.19 (0.84)
($\gamma_0 = 1$)	[0.43, 30.5]	[0.33, 10.5]	[0.43, 192]	[0.30, 8.56]	[0.49, 57.8]
$\hat{\delta}$	0.42 (0.50)	0.57 (0.50)	0.43 (0.50)	0.59 (0.50)	0.42 (0.50)
($\delta_0 = 0.5$)	[0.00, 1.11]	[0.02, 1.31]	[0.00, 1.17]	[0.03, 1.40]	[0.00, 1.10]

Shown: median estimate (median absolute error) [first quartile, third quartile].

Table 6: Monte-Carlo properties of the SEL estimator

Table 6 reports medians, median absolute errors, and quartiles of the estimate distribution by start. The estimates of the exponent δ_0 – arguably the key parameter – vary substantially across runs. Nevertheless, the median estimate is close to 0.5 for the OLS (0.57) and SEuL (0.43) starts; zero and true initialisations are only marginally different. Because δ is constrained to $[0,3]$, the MAE for $\hat{\delta}$ is similar across starts. The slope coefficient on X_1 is stable: median $\hat{\beta}$ is 1.00–1.01 with nearly identical quartiles, reflecting that X_1 and X_2 are IID and that mis-specification in (γ_0, δ_0) has little impact on $\hat{\beta}$. By contrast, the accuracy of $\hat{\gamma}$ is tightly linked to that of $\hat{\delta}$ and exhibits broad dispersion: e g. with the zero start, 25% of $\hat{\gamma}$ exceed 30.5; even with oracle start, 25% exceed 57.8. The intercept $\hat{\alpha}$ shows the most variation across starts.

We conclude that ExEL is an effective safeguard for SEL when convex-hull violations at initial values would otherwise prevent optimisation. Oracle starts are not uniformly superior to simple starts (such as zero); preliminary estimators that acknowledge nonlinearity (SEuL, NLS) are not uniformly better than OLS or zero. In non-linear problems we recommend multi-start strategies in a plausible region or, when feasible, short stochastic searches (such as particle swarm or differential evolution) preceding deterministic optimisation.

6 Conclusion

This paper develops extrapolated empirical likelihood (ExEL) – two simple, computable extensions of the empirical likelihood ratio that remain finite outside the convex hull while leaving the statistic unchanged on a large part of the interior.

The key idea is radial reduction: along any direction through the sample mean, the profile $-2\log\mathrm{ELR}$ for the mean behaves like a one-dimensional strictly convex function, so we splice its outer branch either to a local quadratic (ExEL1, via a second-order Taylor expansion) or to the directional Wald parabola (ExEL2) using a convex bridge. This preserves the desirable interior geometry of EL, restores numerical well-posedness when the convex-hull condition fails, and makes bootstrap calibration feasible on every resample without altering the statistic where EL is already valid.

On the practical side, ExEL acts as a robust 'guardrail'. In small samples and under heavy tails – where convex-hull failures are common – ExEL delivers a finite objective for every data set and for every bootstrap resample, enabling (i) percentile bootstrap calibration, (ii) bootstrap estimation of the Bartlett factor, and (iii) smooth optimisation in problems such as smoothed empirical likelihood for non-linear regression. Simulations across light- and heavy-tailed designs show that plain χ^2 calibration under-covers at small n; moment-based Bartlett correction is a fast, effective default at moderate n; and ExEL coupled with percentile bootstrap delivers the most reliable coverage in the 'hard' regimes (small n, skew/heavy tails) while matching EL in 'easy' ones. The SEL illustration further highlights that ExEL prevents optimisation breakdowns caused by local EL being $-\infty$ at common starting values, without changing any inner solution when the EL constraints are feasible.

The present approach is not limited to EL and can be applied to any situation in which there is an objective function with a convex-hull-like constraint. In particular, it may be used to extrapolate or regularise any LR-like statistic or, more generally, any numerically problematic criterion that shoots off to infinity in the region of interest. It is as general as any Taylor-series-based extrapolation and thus may be used with GEL-family objectives as well.

There are also limitations and open questions. For ExEL1 the Taylor splice is radially convex but not globally convex in θ; characterising when level sets remain convex across directions may be of interest. For ExEL2, a full characterisation of the existence and uniqueness of the exponential bridge for numerically problematic samples would be valuable; a condition for convexity of the level sets in the transition zone could also be derived. On the computational side, adaptive rules for selecting splice radii (or target levels) and warm-starts for the directional Newton system merit further study. For bootstrap Bartlett, stability under severe skew suggests that robust estimators (trimmed/Huberised moments or robust t-bootstrap variants) are worth formal analysis.

In summary, ExEL keeps the strengths of EL where it already works and supplies a convex extension where EL is undefined – thereby enabling uniform computation and calibration without distorting the interior statistic. We view ExEL as a practical complement to existing domain-fixing approaches (e. g. adjusted/extended EL) and a useful building block for moment-based inference, especially in small, irregular, or weighted samples. We hope that the simple radial construction and the empirical guidance here stimulate further theoretical work and broader adoption in applications.

References

Ai, C., & Chen, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, *71*(6), 1795–1843. https://doi.org/10.1111/1468-0262.00470

Antoine, B., Bonnal, H., & Renault, E. (2007). On the efficient use of the informational content of estimating equations: Implied probabilities and Euclidean empirical likelihood. *Journal of Econometrics*, *138*(2), 461–487. https://doi.org/10.1016/j.jeconom.2006.05.005

Baggerly, K. (1998). Empirical likelihood as a goodness-of-fit measure. *Biometrika*, *85*(3), 535–547. https://doi.org/10.1093/biomet/85.3.535

Baragona, R., Battaglia, F., & Cucina, D. (2015). Empirical likelihood for outlier detection and estimation in autoregressive time series. *Journal of Time Series Analysis*, *37*(3), 315–336. https://doi.org/10.1111/jtsa.12145

Baragona, R., Battaglia, F., & Cucina, D. (2017). Empirical likelihood ratio in penalty form and the convex hull problem. *Statistical Methods & Applications*, *26*(4), 507–529. https://doi.org/10.1007/s10260-017-0382-2

Bartolucci, F. (2007). A penalized version of the empirical likelihood ratio for the population mean. *Statistics & Probability Letters*, *77*(1), 104–110. https://doi.org/10.1016/j.spl.2006.05.016

Bergsma, W., Croon, M., & van der Ark, L. A. (2012). The empty set and zero likelihood problems in maximum empirical likelihood estimation. *Electronic Journal of Statistics*, *6*. https://doi.org/10.1214/12-ejs750

Brown, B. M., & Chen, S. X. (1998). Combined and least squares empirical likelihood. *Annals of the Institute of Statistical Mathematics*, *50*, 697–714. https://doi.org/10.1023/A:1003760813552

Calonico, S., Cattaneo, M. D., & Farrell, M. H. (2018). On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association*, *113*(522), 767–779. https://doi.org/10.1080/01621459.2017.1285776

Chen, J., & Huang, Y. (2013). Finite-sample properties of the adjusted empirical likelihood. *Journal of Nonparametric Statistics*, *25*(1), 147–159. https://doi.org/10.1080/10485252.2012.738906

Chen, J., Variyath, A. M., & Abraham, B. (2008). Adjusted empirical likelihood and its properties. *Journal of Computational and Graphical Statistics*, *17*(2), 426–443. https://doi.org/10.1198/106186008x321068

Chen, S. X., & Cui, H. (2006). On Bartlett correction of empirical likelihood in the presence of nuisance parameters. *Biometrika*, *93*(1), 215–220.

Chen, S. X., & Cui, H. (2007). On the second-order properties of empirical likelihood with moment restrictions. *Journal of Econometrics*, *141*(2), 492–516.

Cosma, A., Kostyrka, A. V., & Tripathi, G. (2026). Missing endogenous variables in conditional moment restriction models. *Working paper DP2024-01 (conditionally accepted to the Journal of Business and Economic Statistics)*. https://hdl.handle.net/10993/60100

Devroye, L. P., & Wagner, T. J. (1977). The strong uniform consistency of nearest neighbor density estimates. *The Annals of Statistics*, *5*(3), 536–540. https://doi.org/10.1214/aos/1176343851

DiCiccio, T. J., Hall, P., & Romano, J. (1991). Empirical likelihood is Bartlett-correctable. *Annals of Statisics*, *19*(2), 1053–1061. https://doi.org/10.1214/aos/1176348137

Emerson, S. C., & Owen, A. B. (2009). Calibration of the empirical likelihood method for a vector mean. *Electronic Journal of Statistics*, *3*, 1161–1192. https://doi.org/10.1214/09-ejs518

Grendár, M., & Judge, G. (2009). Empty set problem of maximum empirical likelihood methods. *Electronic Journal of Statistics*, *3*, 1542–1555. https://doi.org/10.1214/09-EJS528

Kitamura, Y. (2007). Empirical likelihood methods in econometrics: Theory and practice. In R. Blundell, W. Newey & T. Persson (Eds.), *Advances in economics and econometrics* (pp. 174–237, Vol. 3). Cambridge University Press. https://doi.org/10.1017/cbo9780511607547.008

Kitamura, Y., Tripathi, G., & Ahn, H. (2004). Empirical likelihood based inference in conditional moment restriction models. *Econometrica*, *72*, 1667–1714. https://doi.org/10.1111/j.1468-0262.2004.00550.x

Lahiri, S. N., & Mukhopadhyay, S. (2012). A penalized empirical likelihood method in high dimensions. *The Annals of Statistics*, *40*(5). https://doi.org/10.1214/12-aos1040

Liu, P., & Zhao, Y. (2022). A review of recent advances in empirical likelihood. *WIREs Computational Statistics*, *15*(3). https://doi.org/10.1002/wics.1599

Liu, Y., & Chen, J. (2010). Adjusted empirical likelihood with high-order precision. *The Annals of Statistics*, *38*(3). https://doi.org/10.1214/09-aos750

Newey, W. K., & McFadden, D. (1994). Large sample estimation and hypothesis testing. In R. F. Engle & D. L. McFadden (Eds.), *Handbook of econometrics, vol. IV* (pp. 2111–2245). Elsevier, The Netherlands.

Newey, W. K., & Smith, R. J. (2004). Higher-order properties of GMM and generalized empirical likelihood estimators. *Econometrica*, *72*, 219–255. https://doi.org/10.1111/j.1468-0262.2004.00482.x

Nguyen, M. K., Phelps, S., & Ng, W. L. (2015). Simulation based calibration using extended balanced augmented empirical likelihood. *Statistics and Computing*, *25*(6), 1093–1112. https://doi.org/10.1007/s11222-014-9506-9

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, *75*, 237–249. https://doi.org/10.1093/biomet/75.2.237

Owen, A. B. (2001). *Empirical likelihood*. Chapman & Hall / CRC, New York. https://doi.org/10.1201/9781420036152

Owen, A. B. (2013). Self-concordance for empirical likelihood. *Canadian Journal of Statistics*, *41*(3), 387–397. https://doi.org/10.1002/cjs.11183

Owen, A. B. (2017). *A weighted self-concordant optimization for empirical likelihood* (tech. rep.). Retrieved August 30, 2023, from https://artowen.su.domains/empirical/countnotes.pdf

Qin, J., & Lawless, J. (1994). Empirical likelihood and general estimating equations. *Annals of Statistics*, *22*, 300–325. https://doi.org/10.1214/aos/1176325370

Smith, R. J. (2007). Efficient information theoretic inference for conditional moment restrictions. *Journal of Econometrics*, *138*(2), 430–460. https://doi.org/10.1016/j.jeconom.2006.05.004

Tsao, M. (2004). Bounds on coverage probabilities of the empirical likelihood ratio confidence regions. *The Annals of Statistics*, *32*(3), 1215–1221. https://doi.org/10.1214/009053604000000337

Tsao, M. (2013). Extending the empirical likelihood by domain expansion. *Canadian Journal of Statistics*, *41*(2), 257–274. https://doi.org/10.1002/cjs.11175

Tsao, M., & Wu, F. (2013). Empirical likelihood on the full parameter space. *The Annals of Statistics*, *41*(4). https://doi.org/10.1214/13-aos1143

Tsao, M., & Wu, F. (2014). Extended empirical likelihood for estimating equations. *Biometrika*, *101*(3), 703–710. https://doi.org/10.1093/biomet/asu014

Zhang, X., & Shao, X. (2016). On the coverage bound problem of empirical likelihood methods for time series. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *78*(2), 395–421. https://doi.org/10.1111/rssb.12119

A Behaviour of ELR

A.1 Convexity of NLELR(θ)

Lemma 2 *$f(\theta) := \mathrm{NLELR}(\theta)$ is strictly convex on the interior of $(Z_{(1)}, Z_{(n)})$ for the uni-variate mean EL with non-degenerate centred data ($\sum_i Z_i = 0, \sum_i Z_i^2 > 0$).* □

PROOF For θ in the convex hull, set $g_i(\theta) := Z_i - \theta$ and $u_i(\theta) := 1 + \lambda(\theta) g_i(\theta)$, where $\lambda(\theta)$ solves

$$h(\lambda(\theta), \theta) := \sum_{i=1}^{n} \frac{Z_i - \theta}{1 + \lambda(\theta)(Z_i - \theta)} = \sum_{i=1}^{n} \frac{g_i(\theta)}{u_i(\theta)} = 0 \quad \text{for } u_i(\theta) > 0 \quad \forall i.$$

Then $\log \mathrm{ELR}(\theta) = -\sum_{i=1}^{n} \log u_i(\theta)$, and $f(\theta) := -2 \log \mathrm{ELR}(\theta)$.

Suppressing the argument θ for brevity, define

$$S_0 := \sum_i 1/u_i, \quad S_1 := \sum_i 1/u_i^2, \quad T_0 := \sum_i g_i/u_i, \quad T_1 := \sum_i g_i/u_i^2, \quad T_2 := \sum_i g_i^2/u_i^2.$$

All are finite on the interior; $S_0, S_1, T_2 > 0$ and $T_0 \equiv 0$.

Using $u_i' = \lambda' g_i - \lambda$, $g_i' = -1$, we get

$$f' = 2 \sum_i \frac{\lambda' g_i - \lambda}{u_i} = 2\lambda' T_0 - 2\lambda S_0 = -2\lambda S_0.$$

because $T_0 \equiv 0$. This implies that f' has the same sign as $-\lambda$.

Differentiate $h(\lambda(\theta), \theta) \equiv 0$ to get $0 = h_\theta + h_\lambda \lambda'$. With $\frac{\partial}{\partial \theta} u(\lambda, \theta) = -\lambda$ and $\frac{\partial}{\partial \lambda} u(\lambda, \theta) = g$, we obtain $h_\theta = -S_0 + \lambda T_1$, $h_\lambda = -T_2$. Since $S_0 - \lambda T_1 = \sum_i (1/u_i - \lambda g_i/u_i^2) = \sum_i (u_i - \lambda g_i)/u_i^2 = \sum_i 1/u_i^2 = S_1$,

$$\lambda' = -\frac{h_\theta}{h_\lambda} = \frac{\lambda T_1 - S_0}{T_2} = -\frac{S_1}{T_2} < 0.$$

Further differentiating $f' = -2\lambda S_0$ yields $f'' = -2\lambda' S_0 - 2\lambda S_0'$. From $u_i' = \lambda' g_i - \lambda$, we have $S_0' = -\lambda' T_1 + \lambda S_1$, and using $S_0 - \lambda T_1 = S_1$ gives

$$f'' = -2\lambda'(S_0 - \lambda T_1) - 2\lambda^2 S_1 = -2\lambda' S_1 - 2\lambda^2 S_1 = 2S_1^2/T_2 - 2\lambda^2 S_1 = 2S_1(S_1 - \lambda^2 T_2)/T_2.$$

Now note that $S_1 - \lambda^2 T_2 = \sum_i (1 - \lambda^2 g_i^2)/u_i^2 = \sum_i (1 - \lambda g_i)(1 + \lambda g_i)/u_i^2 = \sum_i (1 - \lambda g_i)/u_i = S_0 - \lambda T_0 = S_0$. Therefore,

$$f'' = \frac{2S_0 S_1}{T_2} = 2 \frac{\sum_i u_i^{-1} \sum_i u_i^{-2}}{\sum_i g_i^2/u_i^2}.$$

Hence, $f''(\theta) > 0$ for all $\theta \in (0, Z_{(n)})$, and f is strictly convex on that interval. ■

Note that if θ approaches $Z_{(n)}$, at least one $u_i \to 0$, causing $f''(\theta)$ to explode.

Corollary 2 *At $\theta = 0$, $\lambda(\theta) = 0$, $u_i \equiv 1$, $g_i = Z_i$, therefore, $f''(0) = 2n^2/\sum Z_i^2$ (for centred Z_i), i. e. twice the inverse plug-in variance of the sample mean, $2/\mathrm{Var}\, \bar{Z}_n$.* □

A.2 Behaviour of $f(\theta)$ around $\theta = 0$

Using the earlier result that $f'(\theta) = -2\lambda(\theta) \sum_i (1 + \lambda(\theta)(Z_i - \theta))^{-1}$, we study the local behaviour of λ at zero. Let $M_k := \sum_i Z_i^k$ (for centred Z_i), e. g. $M_2 := \sum_i Z_i^2$. The partial derivatives of $h(\lambda(\theta), \theta) = \sum_{i=1}^{n} (Z_i - \theta)/(1 + \lambda(\theta)(Z_i - \theta))$ are equal to

$$\partial_\lambda h = -\sum_i \frac{(Z_i - \theta)^2}{(1 + \lambda(Z_i - \theta))^2} \quad \text{and} \quad \partial_\theta h = -S_0 + \lambda T_1,$$

which, evaluated at $(\lambda,\theta)=(0,0)$, yields $\partial h_\lambda(0,0)=-M_2$ and $\partial_\theta h(0,0)=-n$. Hence,

$$\lambda'(0)=-\left.\frac{\partial_\theta h}{\partial_\lambda h}\right|_{(0,0)}=-\frac{n}{M_2}<0.$$

With $g_i(\theta)=Z_i-\theta$ and centred Z_i, we have $\sum g_i=-n\theta$, $\sum g_i^2=M_2+n\theta^2$, $\sum g_i^3=M_3-3\theta M_2-n\theta^3$, and $\sum g_i^4=M_4-4\theta M_3+6\theta^2M_2+n\theta^4$. Expanding $f(\theta)$ is identical to expanding $2\sum_i\log u_i$, which requires Taylor series of $\lambda(\theta)$ and $f(\theta)$. We rely on the series expansions

$$\frac{1}{1+\lambda g_i}=1-\lambda g_i+\lambda^2g_i^2-\lambda^3g_i^3+O(\lambda^4g_i^4)\quad\text{and}\quad\log(1+\lambda g_i)=\lambda g_i-\frac{\lambda^2g_i^2}{2}+\frac{\lambda^3g_i^3}{3}+O(\lambda^4g_i^4).$$

Near $\theta=0$, $\lambda(\theta)=O(\theta)$, as we show further. Assume a power series (without intercept because $\lambda(0)=0$):

$$\lambda(\theta)=a_1\theta+a_2\theta^2+a_3\theta^3+O(\theta^4).$$

Then $\lambda^2(\theta)=a_1^2\theta^2+2a_1a_2\theta^3+O(\theta^4)$ and $\lambda^3(\theta)=a_1^3\theta^3+O(\theta^4)$. When solving for a_1,a_2, we keep only the terms up to $O(\theta^2)$.

Inserting the geometric expansion into $h(\lambda,\theta)=0$ provides

$$0=\sum_i\frac{g_i}{1+\lambda g_i}=\sum_i(g_i-\lambda g_i^2+\lambda^2g_i^3-\lambda^3g_i^4+O(\lambda^4g_i^5)),$$

where the term-by-term breakdown up to $O(\theta^2)$ is as follows.

1. $\sum_i g_i=-n\theta$ contributes only to θ^1;
2. $-\lambda\sum_i g_i^2=-(a_1\theta+a_2\theta^2)(M_2+n\theta^2)=-a_1M_2\theta-a_2M_2\theta^2+O(\theta^3)$;
3. $\lambda^2\sum_i g_i^3=(a_1^2\theta^2+2a_1a_2\theta^3)(M_3-3\theta M_2-n\theta^3)=a_1^2M_3\theta^2+O(\theta^3)$;
4. $-\lambda^3\sum g_i^4$ yields only $O(\theta^3)$ and higher-order terms.

We plug the expressions for the sums, obtaining

$$\begin{aligned}0&=\sum_i g_i-\lambda\sum_i g_i^2+\lambda^2\sum_i g_i^3-\lambda^3\sum_i g_i^4+\dots\\&=(-n-a_1M_2)\theta+(-a_2M_2+a_1^2M_3)\theta^2+O(\theta^3).\end{aligned}$$

Matching coefficients gives $a_1=-n/M_2$, $a_2=a_1^2M_3/M_2=n^2M_3/M_2^3$, hence

$$\lambda(\theta)=-\frac{n}{M_2}\theta+\frac{n^2M_3}{M_2^3}\theta^2+O(\theta^3).$$

Now expand

$$f(\theta)=2\sum_i\log(1+\lambda g_i)=2\sum_i(\lambda g_i-\lambda^2g_i^2/2+\lambda^3g_i^3/3)+O(\theta^4)=2\lambda\sum_i g_i-\lambda^2\sum g_i^2+\frac{2}{3}\lambda^3\sum g_i^3+O(\theta^4).$$

Substituting the series and keeping terms up to $O(\theta^3)$ gives

$$2\lambda\sum_i g_i=2(a_1\theta+a_2\theta^2)(-n\theta)+O(\theta^4)=-2na_1\theta^2-2na_2\theta^3+O(\theta^4),$$

$$-\lambda^2\sum g_i^2=-(a_1^2\theta^2+2a_1a_2\theta^3)(M_2+n\theta^2)+O(\theta^4)=-a_1^2M_2\theta^2-2a_1a_2M_2\theta^3+O(\theta^4),$$

$$\frac{2}{3}\lambda^3\sum_i g_i^3=\frac{2}{3}(a_1^3\theta^3)(M_3-3\theta M_2-n\theta^3)=\frac{2}{3}a_1^3M_3\theta^3+O(\theta^4).$$

Adding the three contributions yields the coefficient on θ^2 equal to $-2na_1 - a_1^2 M_2 = -2n(-n/M_2) - (n^2/M_2^2)M_2 = n^2/M_2$, on θ^3 equal to $-2na_2 - 2a_1 a_2 M_2 + \frac{2}{3}a_1^3 M_3$, which simplifies to $\frac{2}{3}a_1^3 M_3 = -\frac{2}{3}\frac{n^3 M_3}{M_2^3}$ because $n + a_1 M_2 = 0$. This yields

$$f(\theta) = \frac{n^2}{M_2}\theta^2 - \frac{2}{3}\frac{n^3 M_3}{M_2^3}\theta^3 + O(\theta^4). \tag{A.1}$$

Thus the leading term is quadratic with curvature $f''(0) = 2n^2/M_2$ (twice the inverse of the plug-in variance of $\bar{Z}_n$). The cubic term depends on skewness via M_3; hence, it vanishes for symmetric data. Finally, $f'(\theta)/\theta = f''(0) - 2\frac{n^3 M_3}{M_2^3}\theta + O(\theta^2)$, so $r(\theta) := f'(\theta)/\theta$ is nearly constant near 0, with initial slope governed by M_3; it need not be globally monotone.

B Splice derivation

Let $f(\cdot) \in \mathscr{C}^2$ be the strictly convex (as established in Appendix A.1) NLELR with a vertical asymptote at $Z_{(n)} > 0$, to be spliced to create a smooth transition to a zero-centred quadratic $W(\theta) = a\theta^2$ ($a > 0$). With data re-centred by $\bar{Z}_n$, we have $f(0) = W(0) = 0$ and $f'(0) = W'(0) = 0$.

At a cut-off point θ_1, either $f(\theta_1) > W(\theta_1)$ (EL dominates) or $f(\theta_1) \le W(\theta_1)$ (Wald dominates); the bridge is chosen accordingly.

B.1 Transition to a lower parabola

We require a convex bridge that leaves f at θ_1, meets W at $\theta_2 > \theta_1$ with matching value and slope, and then grows faster than W. A straight line would cross W, so use

$$h(\theta) := a_0 + a_1\theta + a_2 \exp(\theta - \theta_1),$$

which is increasing and strictly convex on the splice if $a_1, a_2 > 0$ because $h''(\theta) = a_2 \exp(\theta - \theta_1) > 0$. Enforce $\mathscr{C}^1$ contact at θ_1 and θ_2; define $t := \theta_2 - \theta_1$ and write

$$\begin{cases} a_0 + a_1\theta_1 + a_2 = f(\theta_1), \\ a_0 + a_1\theta_2 + a_2 t = W(\theta_2), \\ a_1 + a_2 = f'(\theta_1), \\ a_1 + a_2 \exp t = W'(\theta_2). \end{cases}$$

Let $f_1 := f(\theta_1)$, $f_1' := f'(\theta_1)$, $W_2 := W(\theta_2)$, $W_2' := W'(\theta_2) = 2a(\theta_1 + t)$, $t := \theta_2 - \theta_1 > 0$. Solving gives $a_2 = (W_2' - f_1')/(\exp t - 1)$, $a_1 = f_1' - a_2$, and the scalar equation

$$G(t) = f_1' t + \frac{W_2' - f_1'}{\exp t - 1}(\exp t - t - 1) - (W_2 - f_1) = 0,$$

determines $t > 0$ (hence θ_2); then $a_0 = f_1 - a_1\theta_1 - a_2$. Search over $t \ge t_0 := \max\{0, f_1'/(2a) - \theta_1\}$ to ensure $a_2 \ge 0$.

Existence follows from $G(0^+) = f_1 - W(\theta_1) > 0$ and $G(t) \to -\infty$; on $[t_0, \infty)$, G is decreasing in practice, so take the largest root. Uniqueness holds if the root search is restricted to $t \ge t_0$, ensuring $a_2 > 0$; $G'(t) < 0$ for all sufficiently large t.

An example yielding a configuration for this case is the non-centred sample $Z = \{-4, -3, \ldots, 5\}$ with observation multiplicities 1, 2, …, 10; in this case, $\theta_1 = 2.997$, $f(\theta_1) = Q_{\chi_1^2}(0.999) = 10.83$, $\theta_2 = 4.762$.

B.2 Transition to a higher parabola

Now $f(\theta_1) := f_1 < W(\theta_1) = a\theta_1^2$, and the bridge should be nearly linear at θ_2 and touch f at θ_1 without crossing. Use

$$h(\theta) := a_0 + a_1\theta + a_2 \exp(-(\theta - \theta_1))$$

It is also convex: $h'(\theta) = a_1 - a_2 \exp(-(\theta - \theta_1))$, $h''(\theta) = a_2 \exp(-(\theta - \theta_1))$. Impose

$$h(\theta_1) = f_1, \quad h'(\theta_1) = f_1', \quad h(\theta_2) = W(\theta_2) = a\theta_2^2, \quad h'(\theta_2) = W'(\theta_2) = 2a\theta_2.$$

Write $t := \theta_2 - \theta_1 > 0$. Solving the two slope conditions first gives $a_2 = (W_2' - f_1')/(1 - \exp(-t))$, $a_1 = f_1' + a_2$. The single equation

$$G(t) := f_1' t + \frac{W'(\theta_1 + t) - f_1'}{1 - \exp(-t)}(t + \exp(-t) - 1) - (W(\theta_1 + t) - f_1) = 0. \tag{B.1}$$

determines $t > 0$. Finally, $a_0 = f_1 - a_1\theta_1 - a_2$.

Since $1 - \exp(-t) > 0$ for $t > 0$, $\operatorname{sign} a_2 = \operatorname{sign}(2a\theta_2 - f_1')$. Enforcing $t > \max\{0, t_0 := f_1'/(2a)\}$ guarantees $a_2 > 0$ and $h'' > 0$ (convexity) and $h'(\theta_2) = 2a\theta_2 > 0$ (monotonicity at the tangency point).

Here $G(0^+) = f_1 - W(\theta_1) < 0$ and $G(t) \xrightarrow{t\to\infty} +\infty$, so a root exists on $[t_0, +\infty)$. Brackets on $[t_0 + \varepsilon,\ U]$ (small $\varepsilon > 0$), expand U geometrically until G changes sign, and solve with Brent's method.

Example: on the non-centred sample $Z = \{-4, -3, \ldots, 5\}$ with observation multiplicities 10, 9, …, 1, we have $\theta_1 = 0.184$, $f(\theta_1) = Q_{\chi_1^2}(0.999) = 10.83$, and solving $G(\theta_2) = 0$ yields $\theta_2 = 1.667$.

C Derivations for the multi-variate setting

C.1 Behaviour of multi-variate f around zero

With centred data, the EL constraint is

$$h(\lambda, \theta) := \sum_{i=1}^{n} \frac{Z_i - \theta}{1 + \lambda(\theta)'(Z_i - \theta)} = 0.$$

Let $g_i(\theta) := Z_i - \theta$, $u_i(\theta) := 1 + \lambda(\theta)' g_i(\theta)$, and $s_i(\theta) := u_i(\theta) - 1 = \lambda(\theta)' g_i(\theta)$. At $(\lambda, \theta) = (0, 0)$, we have $h(0, 0) = 0$. Now let $\hat{\Sigma} := \frac{1}{n}\sum_i Z_i Z_i'$ be a PD $p \times p$ matrix.

Expand $\lambda(\theta)$ to first order. The partial derivatives at $(0, 0)$ are equal to

$$h_\lambda(0, 0) = -\sum_{i=1}^{n} Z_i Z_i' = -n\hat{\Sigma}, \quad h_\theta(0, 0) = -\sum_{i=1}^{n} I_p = -nI_p,$$

where the latter is because $\partial g_i / \partial \theta = -I_p$ and $u_i \equiv 1$ at $(0, 0)$. For small θ,

$$0 = h_\theta \theta + h_\lambda \lambda(\theta) + O(\|\theta\|^2) \ \Rightarrow\ -n\theta - n\hat{\Sigma}\lambda(\theta) = O(\|\theta\|^2),$$

hence

$$\lambda(\theta) = -\hat{\Sigma}^{-1}\theta + O(\|\theta\|^2).$$

Now expand f to first order.

$$f(\theta) = 2\sum_{i=1}^{n} \log(1 + s_i(\theta)) = 2\sum_{i=1}^{n} (s_i - 0.5 s_i^2) + O\Big(\sum_i |s_i|^3\Big).$$

Summations at leading order use $\sum g_i(\theta) = -n\theta$ and $g_i(\theta) = Z_i + O(\theta)$:

$$\sum_{i=1}^{n} s_i = \lambda' \sum_{i=1}^{n} g_i = \lambda'(-n\theta) = -n\theta'\lambda, \quad \sum_{i=1}^{n} s_i^2 = \lambda' \Big(\sum_{i=1}^{n} g_i g_i' \Big) \lambda = n\lambda' \hat{\Sigma} \lambda + O(\|\theta\|^3).$$

Inserting the expansion for $\lambda(\theta) = -\hat{\Sigma}^{-1}\theta + O(\|\theta\|^2)$ yields

$$\sum_{i=1}^{n} s_i = -n\theta'(-\hat{\Sigma}^{-1}\theta) = n\theta'\hat{\Sigma}^{-1}\theta, \quad \sum_{i=1}^{n} s_i^2 = n(-\hat{\Sigma}^{-1}\theta)'\hat{\Sigma}(-\hat{\Sigma}^{-1}\theta) = n\theta'\hat{\Sigma}^{-1}\theta = \sum_{i=1}^{n} s_i,$$

therefore,

$$f(\theta) = n\theta'\hat{\Sigma}^{-1}\theta + O(\|\theta\|^3). \tag{C.1}$$

This matches the previously derived uni-variate result $f(\theta) = \frac{n^2}{\sum_i Z_i^2}\theta^2 + O(\theta^3)$ (formula (A.1)).

C.2 Behaviour of f_v along a ray around zero

Fix $v \in \mathbb{S}^{p-1}$ and set $\theta = tv$. From Eq. C.1,

$$f_v(t) = nt^2 v'\hat{\Sigma}^{-1} v + O(t^3).$$

Observe that $\mathrm{Var}(\bar{Z}_n v) = n^{-1} v'\hat{\Sigma} v$. In $p > 1$, $n(v'\hat{\Sigma}^{-1} v) \neq 1/(n^{-1} v'\hat{\Sigma} v)$ in general. Since $(v'\hat{\Sigma} v) \cdot (v'\hat{\Sigma}^{-1} v) \geq 1$, we keep the curvature in the form $n(v'\hat{\Sigma}^{-1} v)$.

C.3 Convexity of f

Differentiating the EL constraint and eliminating $D\lambda(\theta)$ yields the Hessian in the form

$$\nabla^2 f(\theta) = 2B' T_2^{-1} B + 2\big(S_1 I_p - T_1 T_1' T_2^{-1}\big),$$

where $B = S_0 I_p - \sum_i u_i(\theta)^{-2}(Z_i - \theta)\lambda(\theta)'$. On the interior, $T_2 > 0$ for a non-degenerate sample, so $B' T_2^{-1} B \geq 0$, and $S_1 I_p - T_1 T_1' T_2^{-1} \geq 0$ by the weighted Cauchy–Schwarz, hence $\nabla^2 f(\theta) > 0$ and f is strictly convex on the interior.

C.4 Derivatives of f_v'

Write

$$g_i(t) := Z_i - \theta_v(t) = Z_i - tv, \quad u_i(t) := 1 + \lambda(t)' g_i(t), \quad f_v(t) := f(\theta_v(t)) = 2\sum_{i=1}^{n} \log u_i(t).$$

Along the ray, $g_i'(t) = -v$, so

$$u_i'(t) = \dot{\lambda}(t)' g_i(t) + \lambda(t)' g_i'(t) = \dot{\lambda}(t)' g_i(t) - \lambda(t)' v,$$

where $\dot{\lambda}(t) := \mathrm{d}\lambda/\mathrm{d}t$. Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}t} f_v(t) = 2\sum_{i=1}^{n} \frac{\dot{\lambda}(t)' g_i(t)}{u_i(t)} - 2(\lambda(t)' v) \sum_{i=1}^{n} \frac{1}{u_i(t)}.$$

By the EL constraint, $\sum_i g_i / u_i = 0$, so the first sum vanishes and we obtain

$$f_v'(t) = -2S_0(t)\lambda(t)' v.$$

Differentiate the constraint $h(\lambda(t), t) = 0$ to get $\dot{\lambda}(t)$:

$$0 = \frac{\mathrm{d}}{\mathrm{d}t} \sum_{i=1}^{n} \frac{g_i}{u_i} = \sum_{i=1}^{n} \Big(\frac{g_i'}{u_i} - \frac{g_i u_i'}{u_i^2} \Big) = -S_0 v - [T_2 \dot{\lambda} - (\lambda' v) T_1],$$

Therefore,

$$\dot{\lambda}(t) = T_2^{-1}[-S_0 v + (\lambda' v) T_1].$$

To obtain the second directional derivative, differentiate $f_v'(t) = -2S_0(t)\lambda(t)' v$ w. r. t. t:

$$f_v''(t) = -2[S_0(t)\dot{\lambda}(t)' v + (\lambda(t)' v) S_0'(t)]$$

Compute the unknown S_0' from $S_0 = \sum_i u_i^{-1}$:

$$S_0'(t) = -\sum_{i=1}^{n} \frac{u_i'}{u_i^2} = -\dot{\lambda} \sum_{i=1}^{n} \frac{g_i}{u_i^2} + (\lambda' v) \sum_{i=1}^{n} u_i^{-2} = -\dot{\lambda}' T_1 + (\lambda' v) S_1.$$

It follows that,

$$f_v''(t) = -2[S_0 \dot{\lambda}' v + (\lambda' v)(-\dot{\lambda}' T_1 + (\lambda' v) S_1)] = -2[\dot{\lambda}'(S_0 v - (\lambda' v) T_1) + (\lambda' v)^2 S_1].$$

Using the shorthand $w_v(t) := S_0 v - (\lambda' v) T_1$, we rewrite the identity from above: $\dot{\lambda}(t) = -T_2^{-1} w_v$. Hence $\dot{\lambda}' w_v = -w_v' T_2^{-1} w_v$. Plugging this back gives the compact second-derivative formula

$$f_v'' = 2 w_v' T_2^{-1} w_v - 2(\lambda' v)^2 S_1.$$

Two remarks help interpret this. First, the term $w_v' T_2^{-1} w_v$ is non-negative and, in case of non-degenerate samples, strictly positive unless we are at the minimum along the ray. Second, $S_1 > 0$, and the terms can be combined to write

$$f_v'' = 2[w_v' T_2^{-1} w_v + (\lambda' v)^2 (T_1' T_2^{-1} T_1 - S_1)].$$

Strict positivity along rays (away from the minimum) is guaranteed by the full Hessian expression; the directional second derivative is $v' \nabla^2 f(\theta_v(t)) v$.

D ExEL with observation multiplicities

EL with observation weights/counts (Owen, 2017) is a popular technique that can reduce computational burden substantially – by more than two orders of magnitude in certain applications where conditioning variables are dummies with many duplicates (Cosma et al., 2026, Appendix B.3). The adaptations are straightforward: the Wald parabola is determined by the weighted variance about the weighted mean, and all derivatives of f simply incorporate the counts w_i inside the sums.

The EL constraint becomes

$$\sum_i \frac{w_i g_i}{u_i} = 0, \quad u_i = 1 + \lambda g_i.$$

Define the weighted aggregates

$$S_0 := \sum_i \frac{w_i}{u_i}, \quad S_1 := \sum_i \frac{w_i}{u_i^2}, \quad T_1 := \sum_i \frac{w_i g_i}{u_i^2}, \quad T_2 := \sum_i \frac{w_i g_i^2}{u_i^2}.$$

Then, exactly as in the unweighted case (with w_i inside the sums), the derivative of NLELR is $f'(\theta) = -2S_0\lambda$, the slope of λ is $\mathrm{d}\lambda/\mathrm{d}\theta = (-S_0 + \lambda T_1)/T_2$, and the second derivative becomes $f''(\theta) = 2\frac{(S_0 - \lambda T_1)^2}{T_2} - 2\lambda^2 S_1$.

For ExEL2 with centred data, centre at the *weighted* sample mean. Let $n_w = \sum_i w_i$, $\bar{Z}_w := \sum_i w_i Z_i / n_w$, and $V_w := \sum_i w_i (Z_i - \bar{Z}_w)^2 / n_w$. Then the variance of the (weighted) sample average is $\mathrm{Var}\, \bar{Z}_w = V_w / n_w$, so the Wald parabola is

$$W(\theta) = a\theta^2, \quad a = n_w / V_w.$$

E On the Bartlett correction

Notation in this section follows Y. Liu and Chen (2010, Section 3). Let Z be a random variable and $g(Z,\theta)_{q\times 1}$ the moment function of dimension q such that $\mathbb{E}g(Z,\theta_0) = 0$ and some technical requirements are satisfied (Y. Liu & Chen, 2010, Theorem 1). Write $V(\theta) := \widehat{\mathrm{Var}}g(Z,\theta)$ (without the $(n-1)$ correction). An eigen-decomposition gives $V(\theta_0) = P\,\mathrm{diag}(\xi_1,\ldots,\xi_q)P'$ with $PP' = I_q$ and $\xi_1,\ldots,\xi_q$ the eigenvalues. Define the rotated moment conditions $Y := P'g(Z,\theta_0)$; then $\mathbb{E}(Y^{(i)})^2 = \xi_i$.

Let $\alpha^{rs\ldots t} := \mathbb{E}Y^{(r)}Y^{(s)}\cdots Y^{(t)}$. Because the rotation diagonalises $V(\theta_0)$, $\alpha^{rr} = \xi_r$ and $\alpha^{rs} = 0$ for $r \neq s$. The Bartlett factor is

$$b = \frac{1}{q}\left(\frac{1}{2}\sum_{r=1}^{q}\sum_{s=1}^{q}\frac{\alpha^{rrss}}{\alpha^{rr}\alpha^{ss}} - \frac{1}{3}\sum_{r=1}^{q}\sum_{s=1}^{q}\sum_{t=1}^{q}\frac{\alpha^{rrstt}}{\alpha^{rr}\alpha^{ss}\alpha^{tt}}\right).$$

As $n\to\infty$, the moment-based estimator using $\hat{\alpha}^{rs\ldots t} := \frac{1}{n}\sum_{i=1}^{n} Y_i^{(r)}Y_i^{(s)}\cdots Y_i^{(t)}$ converges in probability to the population factor. Y. Liu and Chen (2010) propose finite-sample corrections to improve the inferential properties of the Bartlett factor estimator, but these can be insufficient or may compromise positivity of the components of the decomposition of b. An alternative is a bootstrap estimate of the Bartlett factor (S. X. Chen & Cui, 2006). We use trimming to improve robustness:

1. Compute $\hat{\theta}$ in the full sample (for the mean, $\hat{\theta} = \bar{Z}_n$);
2. In each bootstrap replication, evaluate the test statistic at $\hat{\theta}$ using ExEL (thus enforcing the null);
3. With trimming fraction α (e. g. 10% per tail), divide the trimmed mean of the bootstrap -2 ExEL values by the conditional mean of the identically trimmed χ^2_p distribution,

$$(1-2\alpha)^{-1}\int_{Q_{\chi^2_p}(\alpha)}^{Q_{\chi^2_p}(1-\alpha)} x f_{\chi^2_p}(x)\,\mathrm{d}x.$$

The resulting ratio yields $\hat{c} = (1+\hat{b}/n)$; recover $\hat{b}$ accordingly.