Score-type tests for normal mixtures*

Dante Amengual
CEMFI
<amengual@cemfi.es>

Xinyue Bei
Duke University
<xinyue.bei@duke.edu>

Marine Carrasco
Université de Montréal
<marine.carrasco@umontreal.ca>

Enrique Sentana
CEMFI
<sentana@cemfi.es>

December 2022
Revised: November 2023

Abstract

Testing normality against discrete normal mixtures is complex because some parameters turn increasingly underidentified along alternative ways of approaching the null, others are inequality constrained, and several higher-order derivatives become identically 0. These problems make the maximum of the alternative model log-likelihood function numerically unreliable. We propose score-type tests asymptotically equivalent to the likelihood ratio as the largest of two simple intuitive statistics that only require estimation under the null. One novelty of our approach is that we treat symmetrically both ways of writing the null hypothesis without excluding any region of the parameter space. We derive the asymptotic distribution of our tests under the null and sequences of local alternatives. We also show that their asymptotic distribution is the same whether applied to observations or standardized residuals from heteroskedastic regression models. Finally, we study their power in simulations and apply them to the residuals of Mincer earnings functions.

Keywords: Generalized extremum tests, Higher-order identifiability, Likelihood ratio test, Mincer equations.

JEL: C12, C46

*We would like to thank audiences at the annual SCSE Conference (Montreal, May 2022), the CIREQ Conference in honor of Éric Renault (Montreal, May 2022), the NBER - NSF Time Series Conference (Boston, September 2022), the Women in Econometrics Conference (Toronto, October 2022), the annual CIREQ Econometrics Conference (Montreal, May 2023), the X ICEEE Conference (Cagliari, May 2023), the annual IAAE Conference (Oslo, June 2023), and Vanderbilt University for useful comments and suggestions. We are also grateful to the editor (Serena Ng), associate editor, and two anonymous referees for very useful feedback, as well as to Jiahua Chen and Pengfei Li for providing us with the code to implement their test. Of course, the usual caveat applies. Amengual and Sentana gratefully acknowledge financial support from the Spanish Ministry of Science and Innovation through grant PID2021-128963NB-I00, while Carrasco thanks SSHRC for partial funding.

1 Introduction

Finite mixture distributions arise naturally when an observed population contains two or more underlying subpopulations. Starting with the famous Naples bay crab data that Pearson (1894) analyzed, they are often used to model unobserved heterogeneity in many disciplines. Economics examples include duration analysis (Heckman and Singer, 1984), measurement errors (Horowitz and Manski, 1995), schooling and career choice (Keane and Wolpin, 1997), industrial organization (Berry, Carnall and Spiller, 2006), and multiple equilibria in discrete games (Berry and Tamer, 2006); see Compiani and Kitamura (2016) for a more thorough list of references. They have also been used in other fields, such as finance, where the objective is to capture the observed skewness and kurtosis of asset returns that may result from different market conditions, as well as for identifying "convergence clubs" of countries based on per capita GDP, and within-country clustering in household income and wealth.

In this paper, we focus on finite Gaussian mixtures, which are the most popular, with the seemingly modest objective of testing a normal distribution against a mixture of two normals using *i.i.d.* data. Specifically, suppose that individual observations y_i $(i = 1, \ldots, n)$ can be of two types, each following a normal distribution with mean μ_j and variance σ_j^2, $j = 1, 2$. Crucially, these types are not observed by the econometrician, so from her perspective the probability density function (pdf) of an observation is given by the following linear combination of the pdfs of the two types

$$\lambda\phi\left(\frac{y_i - \mu_1}{\sigma_1}\right) + (1 - \lambda)\,\phi\left(\frac{y_i - \mu_2}{\sigma_2}\right),$$

where ϕ denotes the standard normal pdf.

Studying classical tests of normality against a mixture of two normals is a devilish problem. First, the null hypothesis can be written in two ways: either as $H_0 : \mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2$, or as $H_0 : \lambda\,(1 - \lambda) = 0$. Many papers focus only on one of these two null hypotheses but we treat both together. Another difficulty is linked to the fact that some parameters are not identified under normality, although their identity depends on the path along which one approaches the null. Moreover, when testing $H_0 : \lambda\,(1 - \lambda) = 0$, λ is on the boundary of the parameter space, so standard asymptotic theory no longer applies (see Andrews (2001)). Finally, some parameters are only identified – if at all – through higher-order derivatives (cf. Dovonon and Renault (2013)), which means that studying the properties of the likelihood ratio (LR) tests requires up to an eighth-order expansion. All these aspects make likelihood-based testing for normal mixtures highly nonstandard.

Previous papers investigating the properties of the LR tests in this context include Ghosh and Sen (1985), Hathaway (1985), Chen and Chen (2001), Chen, Chen and Kalbfleisch (2004), Azaïs, Gassiat and Mercadier (2006), and Chen, Ponomareva and Tamer (2014). The closest papers to ours are Chen and Li (2009) and Kasahara and Shimotsu (2015). The main difference is that they only focus on the null $H_0 : \mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2$, while we simultaneously deal with the second null hypothesis $H_0 : \lambda (1 - \lambda) = 0$. Our work is also closely related to Cho and White (2007), who consider both null hypotheses but exclude some corner regions of the parameter space. In this respect, one important contribution of our paper is that we explicitly consider all possible values of the parameters thanks to a novel convenient bijective reparametrization.

To circumvent the unusual features of the LR test, which not only make inference complex but also render the maximum of the log-likelihood function of the alternative model numerically unreliable when the null is true, some authors have proposed moment-based tests. Such an approach goes back to the smooth tests in Neyman (1937). In particular, Quandt and Ramsey (1978) use influence functions derived from the moment generating function, while others compare the empirical characteristic function to the theoretical one under normality (see Amengual, Carrasco and Sentana (2020)), or simply a handful of higher-order moments of the normal distribution, as in Jarque and Bera (1980), Bai and Ng (2005), and Bontemps and Meddahi (2005), who, like Kiefer and Salmon (1983), look at the expected values of Hermite polynomials rather than simple powers.[1]

In this paper, we propose score-type tests based on expansions of the log-likelihood function for three null hypotheses: equality of means and variances, equality of means only, and equality of variances only. In all three cases, we derive their limiting distributions and show that they are asymptotically equivalent to the corresponding LR tests under the null and sequences of local alternatives. At the same time, our statistics are much simpler to compute because they do not require the estimation of the full model under the alternative, with the unknown mean and variance parameters simply replaced by their sample analogs under the null hypothesis. Moreover, they do not require any tuning parameters, unlike the EM tests of Chen and Li (2009) and Kasahara and Shimotsu (2015). Interestingly, when testing for the equality of means and variances, our test statistic coincides with the popular Jarque and Bera's formula involving the sample skewness and kurtosis coefficients, which implies that their moment test is equivalent to the LR test in that context. However, when we look at the global LR test, which explicitly considers the two different ways of writing the null hypothesis, the equivalence disappears.

Empirical researchers in economics and finance, though, are often interested in testing the

[1] Bai and Ng (2001) propose a test for conditional symmetry in time series contexts based on the empirical distribution function that can be used to test the null of normality too; see also Dufour et al (1998) for a comparison of the small-sample properties of various normality tests.

normality of the standardized residuals of an econometric model. For that reason, we investigate if our testing procedure is robust to parameter uncertainty. Importantly, we show that when the mean and variance of the observed variable given some conditioning variables are parametric functions of those variables, replacing the unknown parameters by a Gaussian maximum likelihood estimator obtained under the null does not alter the expressions for our proposed test statistics or their asymptotic properties.

The rest of the paper is organized as follows. In Section 2, we introduce the model and the three null hypotheses. Then, we derive the test statistics and their distributions under both the null and suitable sequences of local alternatives in Section 3, and establish their robustness to parameter uncertainty in Section 4. Next, we discuss the results of our simulation experiments in Section 5, and present an empirical application to Mincer earnings functions in Section 6. Finally, Section 7 concludes, with the proofs collected in the Appendix. Moreover, a supplemental appendix includes extra details for some of those proofs together with other auxiliary results.

2 Model, hypotheses, and overview of the test

We consider the model

$$y = \mu(x, \alpha) + \sigma(x, \alpha)\varepsilon, \tag{1}$$

where μ and σ are known functions of x and a finite-dimensional vector of unknown parameters α, and ε conditional on x is *i.i.d.* with zero mean and unit variance. Observations are given by (x_i, y_i), $i = 1, 2, \ldots, n$, where x_i could be the lagged value of y_i in time-series models, in which case we would assume that ε_i conditional on the past is *i.i.d.* We want to test ε is standard normal against the alternative that it follows a standardized mixture of two normals. As we will show in Section 4, estimation of α does not affect the properties of the test, so for the time being we assume α is known and focus on the case without conditioning variables.

Assuming without loss of generality that $\mu(x_i, \alpha) = 0$ and $\sigma(x_i, \alpha) = 1$, we want to test:
H_0 : y has density $\phi(y)$ against
H_1 : y has density $\lambda\phi\left(\frac{y-\mu_1^*}{\sigma_1^*}\right) + (1-\lambda)\phi\left(\frac{y-\mu_2^*}{\sigma_2^*}\right)$, where

$$\mu_1^* = \frac{\delta(1-\lambda)}{\sqrt{1+\lambda(1-\lambda)\delta^2}}, \quad \mu_2^* = -\frac{\lambda}{1-\lambda}\mu_1^*,$$

$$\sigma_1^{*2} = \frac{1}{\left[1+\lambda(1-\lambda)\delta^2\right]\left[\lambda+(1-\lambda)\exp(\varkappa)\right]} \quad \text{and} \quad \sigma_2^{*2} = \exp(\varkappa)\sigma_1^{*2}, \tag{2}$$

with δ, $\varkappa$, and λ becoming the unknown shape parameters. This parametrization guarantees that the marginal distribution of y has zero-mean and unit-variance both under the null and the alternative. As the labels of the two regimes are not identified, we set $\lambda \geq 1/2$ henceforth.[2]

Let $\vartheta = (\delta, \varkappa, \lambda)$, with $\vartheta \in [-\bar{\delta}, \bar{\delta}] \times [-\overline{\varkappa}, \overline{\varkappa}] \times [1/2, 1]$, with $\overline{\varkappa} < \ln 2$ for reasons that will become clearer in section 3.2.1.[3] We consider three different parameter spaces

$$\begin{aligned} \Theta_1' &= [-\bar{\delta}, \bar{\delta}] \times [-\bar{\varkappa}, \bar{\varkappa}] \times [1/2, 1]\,, \\ \Theta_2' &= [-\bar{\delta}, \bar{\delta}] \times \{0\} \times [1/2, 1] \quad \text{and} \\ \Theta_3' &= \{0\} \times [-\bar{\varkappa}, \bar{\varkappa}] \times [1/2, 1]\,. \end{aligned}$$

Θ_1' corresponds to the case where δ, $\varkappa$, and λ are free to take any values within their respective intervals. In turn, Θ_2' corresponds to the case where $\varkappa$ is constrained to be equal to zero, which is relevant when the econometrician knows that the variance is the same in both regimes. Finally, Θ_3' corresponds to the case where δ is constrained to be equal to zero, which captures the knowledge that the means are zero in both regimes.

As we mentioned in the introduction, normality can be achieved within Θ_1' with $\delta = \varkappa = 0$ or with $\lambda = 1$. When $\delta = \varkappa = 0$, λ is not identified. In turn, when $\lambda = 1$, δ and $\varkappa$ are not identified. The existing literature circumvents these problems by testing either $H_{01} : (\delta, \varkappa) = 0$ for $\lambda \leq 1-\varepsilon < 1$, or $H_{02} : \lambda = 1$ with $\min\{|\delta|, |\varkappa|\} \geq \varepsilon > 0$ (see, e.g., Cho and White (2007), and Kasahara and Shimotsu (2015), among others), which effectively excludes the regions depicted in Figures 1a and 1b, respectively, from the parameter space. However, the "corner case" $\{(\delta, \kappa, \lambda) : \min\{|\delta|, |\kappa|\} < \varepsilon,\ \lambda > 1 - \varepsilon\}$ in Figure 1c will remain off-limits even after combining the admissible parameters in Figures 1a and 1b, and it is not obvious that a test statistic that excludes that cube is asymptotically equivalent to the unrestricted LR test.

Moreover, it is well known that the information matrix of the maximum likelihood estimators of $(\delta, \varkappa)$ is singular under H_0 regardless of the value of λ. To confine those singularities to specific parameters whose first-order derivatives become exactly equal to zero under the null, we use a trick analogous to Kasahara and Shimotsu (2015) and replace $\varkappa$ with

$$\kappa = \varkappa + (2\lambda - 1)\delta^2/3, \tag{3}$$

so that the parameter vector becomes $\theta = (\delta, \kappa, \lambda)$. The most general version of the null hypothesis H_0 can thus be written as either $\lambda = 1$ or $(\delta, \kappa) = 0$. Let

$$\Theta_j = \{(\delta, \kappa, \lambda) : (\delta, \kappa - (2\lambda - 1)\delta^2/3, \lambda) \in \Theta_j'\}, \;\; j = 1, 2, 3. \tag{4}$$

[2] We could label the two components for $\lambda = 1/2$ based on the sign of κ, and if that also failed, we could eventually rely on the sign of δ.

[3] Note that the values of σ_1^{*2} and σ_2^{*2} are bounded away from zero by the definition of the parameter spaces Θ_j', $j = 1, 2, 3$, which rules out poles in the likelihood when the values of σ_1^{*2} or σ_2^{*2} go to zero.

Figures 2a-2d describe Θ_1, as well as the parameter combinations that lead to Gaussianity. In this context, the goal of our paper is to construct a score-type test for each of the three hypotheses mentioned above that is asymptotically equivalent to the analogous LR statistic

$$LR_j = 2\left[\sup_{\theta\in\Theta_j} L_n(\delta,\kappa,\lambda) - L_n(0,0,1)\right] \quad \text{with} \quad L_n(\delta,\kappa,\lambda) = \sum_{i=1}^{n} l_i(\delta,\kappa,\lambda), \tag{5}$$

where l_i is the log-likelihood of y_i given θ, $\mu = 0$ and $\sigma^2 = 1$.[4]

To avoid excluding any region of the relevant parameter spaces, we partition them as follows,

$$\mathcal{P}_{a,j} = \{(\delta,\kappa,\lambda) \in \Theta_j : \max\{|\delta|,|\kappa|\} \le 1-\lambda\} \text{ and } \mathcal{P}_{b,j} = \{(\delta,\kappa,\lambda) \in \Theta_j : \max\{|\delta|,|\kappa|\} \ge 1-\lambda\}$$

for $j = 1, 2, 3$. Testing $H_{0a,j} : \delta = \kappa = 0$ with $\theta \in P_{a,j}$ tests whether one or both of the first two moments of the regimes coincide. In turn, testing $H_{0b,j} : \lambda = 1$ with $\theta \in P_{b,j}$ checks that all the observations come from a single regime. We can thus associate each of the two ways of writing the null hypothesis to one and only one of these parameter subspaces, as illustrated in Figures 2e-2g. To the best of our knowledge, this has never been done before.

3 Test statistics

In this section, we first treat $H_{0a,j}$ and $H_{0b,j}$ separately and develop the corresponding test statistics, which we then combine by taking the largest of the two.

3.1 Inside the pyramid: testing H_{0a}

As we mentioned above, testing $H_{0a,1} : \delta = \kappa = 0$ with $\theta \in P_{a,1}$ assesses whether the mean and variance are the same in both regimes. Similarly, testing $H_{0a,2} : \delta = 0$ with $\theta \in P_{a,2}$ implicitly assumes that the variances are known ex-ante to be the same in both regimes and one simply wants to test whether the mean is also the same. Finally, testing $H_{0a,3} : \kappa = 0$ with $\theta \in P_{a,3}$ maintains that the means of the two regimes are known ex-ante to be 0 and one only wants to check that the variances coincide too.

Let $LR_{a,j}$ be the LR statistics for testing $H_{0a,j}$ with $\theta \in P_{a,j}$, as in Figure 2f, which is given by

$$LR_{a,j} = 2\left[\sup_{\theta\in\mathcal{P}_{a,j}} L_n(\theta) - L_n(0,0,1)\right].$$

Our reparametrization allows us to write the derivatives of the log-likelihood with respect to δ and κ at the point $(0, 0, \lambda)$ using the Hermite polynomials $h_3 = y^3 - 3y$ and $h_4 = y^4 - 6y_2 + 3$

[4] For notational simplicity, we systematically use $L_n(0, 0, 1)$ to denote the log-likelihood function under H_0 even though it coincides with both $L_n(\delta, \kappa, 1)$ and $L_n(0, 0, \lambda)$, avoiding the dependence on μ and σ^2 if unnecessary.

as:

$$\frac{\partial l_i}{\partial \delta} = 0, \quad \frac{\partial l_i}{\partial \kappa} = 0,$$

$$\frac{\partial^2 l_i}{\partial \delta^2} = 0, \quad \frac{\partial^2 l_i}{\partial \delta \partial \kappa} = -\frac{1}{2}\lambda(1-\lambda)h_{3i}, \quad \frac{\partial^2 l_i}{\partial \kappa^2} = \frac{1}{4}\lambda(1-\lambda)h_{4i},$$

$$\frac{\partial^3 l_i}{\partial \delta^3} = 0 \text{ and } \frac{\partial^4 l_i}{\partial \delta^4} = -\frac{2}{3}\lambda(1-\lambda)(1-\lambda+\lambda^2)h_{4i}.$$

Hence, an eighth-order expansion of the log-likelihood function immediately implies that our score-type tests will depend on

$$H_{3,n} = \sum_{i=1}^{n} h_{3i} = \sum_{i=1}^{n} y_i(y_i^2 - 3), \quad V_3 = var\,(h_{3,i}) = 6,$$

$$H_{4,n} = \sum_{i=1}^{n} h_{4i} = \sum_{i=1}^{n} (3 - 6y_i^2 + y_i^4) \text{ and } V_4 = var\,(h_{4,i}) = 24.$$

More formally, the score-type test statistics corresponding to the three null hypotheses and their asymptotic distributions are given by the following result:

Proposition 1 *Let*

$$LM_{a,1} = \frac{H_{3,n}^2}{nV_3} + \frac{H_{4,n}^2}{nV_4}, \quad LM_{a,2} = \frac{H_{3,n}^2}{nV_3} + \frac{H_{4,n}^2}{nV_4}\mathbf{1}\,[H_{4,n} < 0] \quad \text{and} \quad LM_{a,3} = \frac{H_{4,n}^2}{nV_4}\mathbf{1}\,[H_{4,n} > 0]$$

denote the score-type test statistics corresponding to the three null hypotheses, where $1[A]$ *denotes the indicator function for event* A. *Then:*
a) For $j = 1, 2, 3$, $LR_{a,j} = LM_{a,j} + o_p(1)$ *under* $H_{0a,j}$.
b) In addition, under H_0,

$$LM_{a,1} \xrightarrow{d} \chi_2^2, \quad LM_{a,2} \xrightarrow{d} \chi_1^2 + \max(0, Z)^2 \quad \text{and} \quad LM_{a,3} \xrightarrow{d} \max(0, Z)^2,$$

where χ_j^2 *denotes a chi-square random variable with* j *degrees of freedom and* Z *is a standard normal independent of* χ_1^2.

In $LM_{a,1}$ we recognize Jarque and Bera's (1980) test statistic, which exploits both the skewness and kurtosis of the data. Not surprisingly, its asymptotic distribution, as well as that of $LR_{a,1}$ by virtue of the asymptotic equivalence in Proposition 1.a, is χ_2^2 under H_0. In contrast, $LM_{a,3}$ exploits the data's potential lepkurtosis only, while $LM_{a,2}$ both its potential skewness and platykurtosis, which explains their partially one-sided nature. Intuitively, unrestricted two-component Gaussian mixtures, such as the one in Figure 3a, can generate the entire admissible range of skewness-kurtosis coefficients. In contrast, two-component mixtures with a common variance can have either positive or negative skewness but they can only be platykurtic in the vicinity of the null, as illustrated in Figure 3b. Finally, scale mixtures of two Gaussians are symmetric and necessarily leptokurtic, as confirmed by Figure 3c.

3.2 Outside the pyramid: testing H_{0b}

We are now concerned with testing $H_{0b,j} : \lambda = 1$ with $\theta \in \mathcal{P}_{b,j}$, as in Figure 2g. As usual, $H_{0b,1}$ corresponds to the case where both the mean and variance can be different across regimes under the alternative, $H_{0b,2}$ to the case where only the mean may differ across regimes, and $H_{0b,3}$ to the case where only the variance is allowed to change. Importantly, we are in the rather unusual setting where the parameter λ is on the boundary of its range $[1/2, 1]$ and the "nuisance" parameters (δ, κ) are not identified under H_0.

The score with respect to λ at the point $(\delta, \kappa, 1)$ is given by

$$\begin{aligned}\frac{\partial l_i}{\partial \lambda} &= \frac{1}{2}\left(3 - e^{\kappa - \frac{\delta^2}{3}}\right) - \frac{1}{\sqrt{e^{\kappa - \frac{\delta^2}{3}}}} \exp\left\{\frac{1}{2}\left[y_i^2 - (y_i + \delta)^2 e^{-\left(\kappa - \frac{\delta^2}{3}\right)}\right]\right\} \\ &\quad -\delta y_i - \left(1 - e^{\kappa - \frac{\delta^2}{3}}\right)\frac{y_i^2}{2} + \frac{\delta^2}{2}(y_i^2 - 1).\end{aligned} \tag{6}$$

As we explained before, $\partial l_i/\partial\lambda$ equals zero when δ and κ are simultaneously 0. For that reason, we first focus on a region where (δ, κ) is kept away from $(0, 0)$, leaving the discussion of the general case for later.

3.2.1 Outside the pyramidion

Let $B = \{(\delta, \kappa, \lambda) \in \mathcal{P}_{b,1} : \sqrt{\delta^2 + \kappa^2} \geq \epsilon\}$ for some $\epsilon > 0$. Henceforth, we refer to the complement of this subset over $\mathcal{P}_{b,1}$ represented in Figure 2h as the "pyramidion".[5]

For a given (δ, κ), let

$$LM_b(\delta, \kappa) = \left\lfloor \frac{\mathcal{G}_n(\delta, \kappa)}{\sqrt{V(\delta, \kappa)}} \right\rfloor_-^2, \tag{7}$$

where $\lfloor\cdot\rfloor_- = \min(0, \cdot)$, $\mathcal{G}_n(\delta, \kappa) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\partial l_i}{\partial \lambda}$, and $V(\delta, \kappa) = g\left(\delta, \kappa - \delta^2/3; \delta, \kappa - \delta^2/3\right)$, with

$$\begin{aligned} g(\delta_1, k_1, \delta_2, k_2) &= \frac{\exp\left[-\frac{\left(\delta_2^2 e^{k_1} + \delta_1^2 e^{k_2}\right)}{2e^{k_1+k_2}}\right]}{\sqrt{e^{k_1} + e^{k_2} - e^{k_1+k_2}}} \exp\left[-\frac{\left(\delta_2 e^{k_1} + \delta_1 e^{k_2}\right)^2}{2e^{k_1+k_2}\left(e^{k_1+k_2} - e^{k_1} - e^{k_2}\right)}\right] \\ &\quad -\frac{1}{2}\left[3 + 2\delta_1\delta_2 + (\delta_1\delta_2)^2 + \delta_1^2(e^{\kappa_2} - 1) + \delta_2^2(e^{\kappa_1} - 1) - e^{\kappa_1} - e^{\kappa_2} + e^{\kappa_1+\kappa_2}\right].\end{aligned} \tag{8}$$

The one-sided nature of (7) reflects the fact that expected value of (6) is negative for alternatives with λ strictly less than 1 and δ and κ not simultaneously 0. Importantly, the fact that $V(\delta, \kappa)$, which is the variance of (6) under the null, becomes unbounded when $\exp\left(\kappa - \delta^2/3\right) \geq 2$ is inconsequential because of our assumption that $\bar{\varkappa} < \ln 2$. In this context, we define our score

[5] A pyramidion is the capstone of an Egyptian pyramid, and therefore pyramidal itself. By relying on Euclidean distances, though, ours is cylindric instead.

test statistic as

$$LM_{b,1}^{B}=\sup_{(\delta,\kappa,1)\in\mathcal{P}_{b,1}\cap B}LM_{b}\left(\delta,\kappa\right) \tag{9}$$

and the corresponding LR test statistic by

$$LR_{b,1}^{B}=2\left[\sup_{\theta\in\mathcal{P}_{b,1}\cap B}L_{n}(\theta)-L_{n}(0,0,1)\right]. \tag{10}$$

We can then show that:

Proposition 2 *1) Under H_0, we have that*

$$\mathcal{G}_{n}(\delta,\kappa)\equiv\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial l_{i}}{\partial\lambda}\Rightarrow G\left(\delta,\kappa\right)$$

over $\mathcal{P}_{b,1}\cap B$, where $G(\delta,\kappa)$ is a Gaussian process indexed by (δ,κ) with $E[G(\delta,\kappa)]=0$ and

$$cov[G\left(\delta_{1},\kappa_{1}\right),G\left(\delta_{2},\kappa_{2}\right)]=g\left(\delta_{1},\kappa_{1}-\frac{\delta_{1}^{2}}{3},\delta_{2},\kappa_{2}-\frac{\delta_{2}^{2}}{3}\right).$$

2) In addition,

$$(a)\quad LM_{b}\left(\delta,\kappa\right)\Rightarrow\left\lfloor\frac{G(\delta,\kappa)}{\sqrt{V\left(\delta,\kappa\right)}}\right\rfloor_{-}^{2}\quad for\ \ (\delta,\kappa,1)\in\mathcal{P}_{b,1}\cap B,$$

$$(b)\quad LM_{b,1}^{B}\xrightarrow{d}\sup_{(\delta,\kappa,1)\in\mathcal{P}_{b,1}\cap B}\left\lfloor\frac{G(\delta,\kappa)}{\sqrt{V\left(\delta,\kappa\right)}}\right\rfloor_{-}^{2}\quad and$$

$$(c)\quad LR_{b,1}^{B}=LM_{b,1}^{B}+o_{p}\left(1\right).$$

Proposition 2 determines the limits of the empirical processes $G_n(\delta,\kappa)$ (score) and $LM_b(\delta,\kappa)$ (test) under H_0, whence one can obtain the asymptotic null distribution of the equivalent statistics $LR_{b,1}^{B}$ and $LM_{b,1}^{B}$.

3.2.2 Inside the pyramidion

In the previous subsection, we restricted (δ,κ) away from 0. Therefore, we still need to obtain a test statistic that remains valid inside the pyramidion when $(\delta,\kappa)\to 0$, in which case both (6) and $\mathcal{G}_n(\delta,\kappa)$ also go to zero for any given sample. The problem is that

$$\left\{\frac{\mathcal{G}_{n}(\delta,\kappa)}{\sqrt{V\left(\delta,\kappa\right)}}:(\delta,\kappa,1)\in\Theta_{1}\backslash\left\{0,0,1\right\}\right\}$$

is not stochastically equicontinuous[6] because for any given sample of size n we can find sequences $(\delta_m,\kappa_{1m})\to 0$ and $(\delta_{2m},\kappa_{2m})\to 0$ such that

[6]See Andrews (1994) for the definitions of stochastic equicontinuity and weak convergence. An example of failure of equicontinuity is given in Supplemental Appendix B.

$$\lim_{m\to\infty}\frac{\mathcal{G}_n\left(\delta_{1m},\kappa_{1m}\right)}{\sqrt{V\left(\delta_{1m},\kappa_{1m}\right)}}\neq\lim_{m\to\infty}\frac{\mathcal{G}_n\left(\delta_{2m},\kappa_{2m}\right)}{\sqrt{V\left(\delta_{2m},\kappa_{2m}\right)}}.$$

Consequently, we cannot directly rely on the functional central limit theorem underlying Proposition 2.

To deal with this problem, we perform yet another one-to-one reparametrization over $\mathcal{P}_{b,1}$ from (δ,κ,λ) to (τ,φ,η) such that $\tau\to 0$ if and only if $(\delta,\kappa)\to 0$ (see equations (18)-(20) in the proof of Proposition 3 for details). Drawing inspiration from Lee and Chesher (1986), we then define

$$\mathcal{G}_n^{'}(\tau,\varphi)=\frac{1}{\tau}\mathcal{G}_n[\delta(\tau,\varphi),\kappa(\tau,\varphi)]$$

and

$$V^{'}(\tau,\varphi)=\frac{1}{\tau^2}V[\delta(\tau,\varphi),\kappa(\tau,\varphi)].$$

This simple trick guarantees that $\lim_{\tau\to 0}\mathcal{G}_n^{'}(\tau,\varphi)$ is well defined, so that we can show that the empirical process $\{\mathcal{G}_n^{\prime}(\tau,\varphi)\}$ converges weakly to a well defined Gaussian limit (see again the proof of Proposition 3 for details). But since

$$\frac{\mathcal{G}_n^{'}(\tau,\varphi)}{V^{'}(\tau,\varphi)}=\frac{\mathcal{G}_n[\delta(\tau,\varphi),\kappa(\tau,\varphi)]}{V[\delta(\tau,\varphi),\kappa(\tau,\varphi)]}\tag{11}$$

for any $\tau\neq 0$, the one-to-one mapping from (δ,κ,λ) to (τ,φ,η) over the whole of $P_{b,1}$, and therefore over both $P_{b,1}\cap B$ and its complement, implies that we can work with $G_n^{'}(\tau,\varphi)$ over the entire $P_{b,1}$ regardless of the partition. Thus, we can define the score-type tests corresponding to $H_{0b,j}$, with $j=1,2,3$, as:

$$LM_{b,1}=\sup_{(\delta,\kappa):(\delta,\kappa,1)\in\Theta_1\setminus\{0,0,1\}}\left\lfloor\frac{\mathcal{G}_n(\delta,\kappa)}{\sqrt{V\left(\delta,\kappa\right)}}\right\rfloor_-^2,$$

$$LM_{b,2}=\sup_{0<|\delta|\leq\bar{\delta}}\left\lfloor\frac{\mathcal{G}_n(\delta,\frac{\delta^2}{3})}{\sqrt{V\left(\delta,\frac{\delta^2}{3}\right)}}\right\rfloor_-^2$$

and

$$LM_{b,3}=\sup_{0<|\kappa|\leq\bar{\varkappa}}\left\lfloor\frac{\mathcal{G}_n(0,\kappa)}{\sqrt{V\left(0,\kappa\right)}}\right\rfloor_-^2,$$

where we have excluded the element $\{0,0,1\}$ because at this point both $\mathcal{G}_n(0,0)$ and $V(0,0)=0$, although it is easy to see that the avoidable discontinuity of (11) at $\tau=0$ could be easily removed by replacing the ratio by 0.

Let us explain the choice of the spaces over which the supremum is taken. When $\lambda=1$,

$(\delta, \kappa, 1) \in \Theta_1$ is equivalent to $\{(\delta,\kappa) : |\delta| \le \bar{\delta}, |\kappa - \delta^2/3| \le \bar{\varkappa}\}$ in view of (4), so that

$$\sup_{(\delta,\kappa,1)\in\Theta_1\backslash\{0,0,1\}} \left\lfloor \frac{\mathcal{G}_n(\delta,\kappa)}{\sqrt{V(\delta,\kappa)}} \right\rfloor_-^2 = \sup_{|\delta|\le\bar{\delta},|\kappa-\delta^2/3|\le\bar{\varkappa},|\delta|+|\kappa|>0} \left\lfloor \frac{\mathcal{G}_n(\delta,\kappa)}{\sqrt{V(\delta,\kappa)}} \right\rfloor_-^2 .$$

Similarly, $(\delta, \kappa, 1) \in \Theta_2$ is equivalent to $\{(\delta,\kappa) : |\delta| \le \bar{\delta}, \kappa = \delta^2/3\}$, while $(\delta,\kappa,1) \in \Theta_3$ is equivalent to $\{(\delta,\kappa) : \delta = 0, |\kappa| \le \bar{\varkappa}\}$.

In this context, the following proposition establishes the equivalence between the LR test and our proposed one:

Proposition 3 *(a) Under H_0, we have*

$$LM_{b,1} \overset{d}{\to} \sup_{|\delta|\le\bar{\delta},|\kappa-\delta^2/3|\le\bar{\varkappa},|\delta|+|\kappa|>0} \left\lfloor \frac{G(\delta,\kappa)}{\sqrt{V(\delta,\kappa)]}} \right\rfloor_-^2 ,$$

$$LM_{b,2} \overset{d}{\to} \sup_{0<|\delta|\le\bar{\delta}} \left\lfloor \frac{G(\delta,\frac{\delta^2}{3})}{\sqrt{V(\delta,\frac{\delta^2}{3})}} \right\rfloor_-^2 \quad \textit{and}$$

$$LM_{b,3} \overset{d}{\to} \sup_{0<|\kappa|\le\bar{\varkappa}} \left\lfloor \frac{G(0,\kappa)}{\sqrt{V(0,\kappa)}} \right\rfloor_-^2 .$$

(b) Moreover,

$$LR_{b,j} = LM_{b,j} + o_p(1),$$

where

$$LR_{b,j} = 2\left[\sup_{\theta\in\mathcal{P}_{b,j}} L_n(\theta) - L_n(0,0,1)\right].$$

Importantly, the results of Proposition 3, which allow one to obtain the common asymptotic distributions of $LM_{b,j}$ and $LR_{b,j}$ under the relevant null hypothesis, are novel because they hold for the whole space P_b, without the need to exclude the corner case in Figure 1c in which the pyramidion lies.

3.3 Combined test of H_0

In the previous two subsections, we have derived the relevant tests statistics over either $\mathcal{P}_{a,j}$ or $\mathcal{P}_{b,j}$, but we really want to test the null of normality of the entire parameter spaces Θ_j, $j = 1, 2, 3$. Given that $\Theta_j = \mathcal{P}_{a,j} \cup \mathcal{P}_{b,j}$, and that the LR test statistic is such that $LR_j = \max(LR_{a,j}, LR_{b,j})$, we define $LM_j = \max(LM_{a,j}, LM_{b,j})$. Using the asymptotically equivalent results in Propositions 1 and 3, it immediately follows that $LR_j = LM_j + o_p(1)$.

Interestingly, we can show that the test statistic in $\mathcal{P}_a$ is no larger than the one in $\mathcal{P}_b$ with probability 1 for Θ_1 and Θ_3, which implies that the corresponding tests can be simplified as follows:

Proposition 4 *Under H_0 and H_1, we have*

$$\max\left(LM_{a,j}, LM_{b,j}\right) = LM_{b,j} \text{ for } j = 1 \text{ and } 3. \tag{12}$$

Therefore, under H_0, we have

$$LR_1 = \sup_{|\delta|\leq\bar{\delta},|\kappa-\delta^2/3|\leq\bar{\varkappa},|\delta|+|\kappa|>0} \left\lfloor \frac{\mathcal{G}_n(\delta,\kappa)}{\sqrt{V(\delta,\kappa)}} \right\rfloor_-^2 + o_p(1) \tag{13}$$

and

$$LR_3 = \sup_{0<|\kappa|\leq\bar{\varkappa}} \left\lfloor \frac{\mathcal{G}_n(0,\kappa)}{\sqrt{V(0,\kappa)}} \right\rfloor_-^2 + o_p(1). \tag{14}$$

In contrast, the test statistic in $\mathcal{P}_{b,2}$ may be either smaller or larger than that in $\mathcal{P}_{a,2}$ with positive probability asymptotically (see Figure 3e in comparison to Figures 3d and 3f). Consequently, our score-type statistic for testing normality against a finite normal mixture with $\theta \in \Theta_2$ will be

$$LM_2 = \max(LM_{a,2}, LM_{b,2}).$$

3.4 Distribution under local alternatives

Given that there are two ways of expressing the null, there are two natural local alternatives to $H_0 : y_i \sim \mathcal{N}(0,1)$, depending on whether $(\delta,\kappa) \to (0,0)$ or λ goes to 1.

First, we consider local alternatives in which λ is kept fixed somewhere in the range $[1/2, 1-\epsilon]$ while (δ,κ) approaches $(0,0)$. Let P_0 be the distribution of $y_1,\ldots,y_n$ under the null hypothesis and P_{θ_n} be the distribution of $y_1,\ldots,y_n$ under local alternatives such that $\lim_{n\to\infty}(w_{1n}, w_{2n}) = (w_1, w_2) \in \mathbb{R}^2$, where

$$\begin{aligned} w_{1n} &= -\frac{1}{2}\sqrt{n}\delta_n\kappa_n \text{ and} \\ w_{2n} &= \sqrt{n}\left(\frac{1}{8}\kappa_n^2 - \frac{1-\lambda+\lambda^2}{36}\delta_n^4\right). \end{aligned}$$

We will denote the corresponding local alternatives as

$$\mathcal{H}_{1n} : \left\{(w_{1n}, w_{2n}) \text{ such that } \lim_{n\to\infty}(w_{1n}, w_{2n}) = (w_1, w_2) \in \mathbb{R}^2\right\}.$$

Somewhat unusually, we can have $w_{1n} = O(1)$ and $w_{2n} = O(1)$ in two different cases:
(a) when $\sqrt{n}\delta_n\kappa_n = O(1)$ and $\sqrt{n}\kappa_n^2 = O(1)$, or
(b) when $\sqrt{n}\delta_n\kappa_n = O(1)$ and $\sqrt{n}\delta_n^4 = O(1)$.

Second, we also consider local alternatives in which δ and κ are fixed and not simultaneously

0 while λ goes to 1 at the usual $\sqrt{n}$ rate. We will denote these other local alternatives as

$$\mathcal{H}_{2n} : \lambda_n = 1 - \frac{\rho}{\sqrt{n}},$$

where ρ is some positive constant.

Let P_{β,λ_n}, with $\beta = (\delta, \kappa)$, denote the probability measure of $y_1, \ldots, y_n$ corresponding to $\mathcal{H}_{2n}$. In addition, let $\chi^2_k(\upsilon)$ denote a non-central chi-square random variable with k degrees of freedom and non-centrality parameter υ. We can then show that:

Proposition 5 *(a) P_{θ_n} is contiguous with respect to P_0.*
(b) For any $(\beta, 1) \in B$, P_{β,λ_n} is contiguous with respect to P_0.
(c) Under the relevant local alternatives, we have

$$LM_{a,1} = \frac{H^2_{3,n}}{nV_3} + \frac{H^2_{4,n}}{nV_4} \xrightarrow{d} \begin{cases} \chi^2_2\left(V_3 w_1^2 + V_4 w_2^2\right) & \text{under } \mathcal{H}_{1n}, \\ \chi^2_2\left(\frac{c_3^2\rho^2}{V_3} + \frac{c_4^2\rho^2}{V_4}\right) & \text{under } \mathcal{H}_{2n}; \end{cases}$$

$$LM^B_{b,1} = \sup_{(\beta,1)\in\mathcal{P}_{b,1}\cap B} LM_n(\beta) \xrightarrow{d} \begin{cases} \sup\limits_{(\beta,1)\in\mathcal{P}_{b,1}\cap B} \min\left\{0, \frac{G(\beta)+c_3w_1+c_4w_2}{\sqrt{V(\beta)}}\right\}^2 & \text{under } \mathcal{H}_{1n}, \\ \sup\limits_{(\beta,1)\in\mathcal{P}_{b,1}\cap B} \min\left\{0, \sqrt{V(\beta)}\left(\frac{G(\beta)}{V(\beta)} - \rho\right)\right\}^2 & \text{under } \mathcal{H}_{2n}; \end{cases}$$

where $G(\beta)$ and $V(\beta)$ are defined at the beginning of section 3.2.1, and

$$\begin{aligned} c_3 &= cov\left(h_{3i}, \frac{\partial l_i}{\partial \lambda}\right) = \delta^3 + 3\delta\left(e^{\kappa - \frac{\delta^2}{3}} - 1\right) \quad \text{and} \\ c_4 &= cov\left(h_{4i}, \frac{\partial l_i}{\partial \lambda}\right) = 6\delta^2\left(1 - e^{\kappa - \frac{\delta^2}{3}}\right) - \delta^4 - 3\left(1 - e^{\kappa - \frac{\delta^2}{3}}\right)^2. \end{aligned}$$

(d) Moreover, $LM_{a,j}$ and $LR_{a,j}$ are asymptotically equivalent under $\mathcal{H}_{1n}$ while $LM^B_{b,j}$ and $LR^B_{b,j}$, which are defined in (9) and (10), respectively, are asymptotically equivalent under $\mathcal{H}_{2n}$.

The following remarks are in order:

Remark 1 An interesting implication of Proposition 5(c) in terms of power under $\mathcal{H}_{1n}$ is the following. We have

$$c_3 w_1 = \left[\delta^3 + 3\delta\left(e^{\kappa - \frac{\delta^2}{3}} - 1\right)\right] w_1 \leq 0,$$

while the sign of

$$c_4 w_2 = \left[6\delta^2\left(1 - e^{\kappa - \frac{\delta^2}{3}}\right) - \delta^4 - 3\left(1 - e^{\kappa - \frac{\delta^2}{3}}\right)^2\right] w_2$$

depends on both the type of local alternative (either $\sqrt{n}\kappa_n^2 = O(1)$ or $\sqrt{n}\delta_n^4 = O(1)$) and the values taken by δ and κ. Since we take a minimum over δ and κ, we can always find values of these parameters such that $c_4 w_2 \leq 0$. Consequently, the expectation of $\partial l_i / \partial \lambda$ is negative and the test $LM^B_{b,j}$ will have nontrivial power against P_{θ_n}. However, if κ_n and δ_n go to zero too fast, so that $w_1 = o_p(1)$ and $w_2 = o_p(1)$, then the tests $LM_{a,1}$ and $LM^B_{b,j}$ will have trivial power.

Remark 2 The $LM^B_{b,j}$ test has nontrivial power against the local alternatives $\mathcal{H}_{2n}$.

Remark 3 It follows from Proposition 5 that the $LM_{a,1}$ test has non trivial power against $\mathcal{H}_{2n}$ provided either $\delta \neq 0$ or $\kappa \neq 0$. On the other hand, if λ were to go to zero faster than $1/\sqrt{n}$, then $LM_{a,1}$ would not have power even if $\delta \neq 0$ and $\kappa \neq 0$.

Remark 4 The asymptotic distribution of $\max(LM_{a,1}, LM^B_{b,j})$ under $\mathcal{H}_{1n}$ and $\mathcal{H}_{2n}$ could in principle be deduced from Proposition 5(c), although there is no simple expression for it.

Remark 5 Proposition 5(d) implies that the asymptotic distribution of $LR_{a,j}$ is the same as that of $LM_{a,j}$ under local alternatives H_{1n}, and the asymptotic distribution of $LR^B_{b,j}$ is the same as that of $LM^B_{b,j}$ under local alternatives H_{2n}.

We would like to emphasize that Proposition 5 implies that our tests are consistent for any fixed alternative for which $\lambda \neq 1$ and either $\delta \neq 0$ or $\kappa \neq 0$. Indeed, the different test statistics diverge under such fixed alternatives, so their power goes to 1. Consequently, the following corollary holds:

Corollary 1 *Under fixed alternatives for which* $\lambda \neq 1$ and either $\delta \neq 0$ or $\kappa \neq 0$, *we have that*

$$LM_{a,1} \to \infty \quad \text{and} \quad LM^B_{b,j} \to \infty$$

as n *goes to infinity.*

4 Robustness to parameter uncertainty

In this section, we study the impact of estimating the mean and variance parameters under the null on the asymptotic properties of our testing procedures. Specifically, we consider the case where the conditional mean and variance of y are parametric functions of another observable variable x, as in (1). Autoregressive and GARCH models are particular examples in which x contains lagged values of y. In this context, the objective becomes to test whether the standardized innovation ε follows a standard normal distribution versus a standardized mixture of two Gaussian components.

The conditional log-likelihood of the i^{th} observation is given by

$$k - \frac{1}{2}\ln\sigma_Y(x_i,\alpha) + \ln\left\{\frac{\lambda}{\sqrt{\sigma_1^{*2}}}\exp\left[-\frac{1}{2\sigma_1^{*2}}\left(\frac{y_i-\mu_Y(x_i,\alpha)}{\sqrt{\sigma_Y^2(x_i,\alpha)}}-\mu_1^*\right)^2\right]\right.$$
$$\left.+\frac{1-\lambda}{\sqrt{\sigma_2^{*2}}}\exp\left[-\frac{1}{2\sigma_2^{*2}}\left(\frac{y_i-\mu_Y(x_i,\alpha)}{\sqrt{\sigma_Y^2(x_i,\alpha)}}-\mu_2^*\right)^2\right]\right\},$$

where k is the constant of integration and μ_1^*, μ_2^*, σ_1^{*2} and σ_2^{*2} are defined in (2).

Assumption 1 *$\mu_Y(x_i, \alpha)$ and $\sigma_Y(x_i, \alpha)$ are eight times continuously differentiable with respect to α.*

Assumption 2 *For all $k \in N^{d_\alpha}$ and $\iota' k = 1, \ldots, 8$, it holds that*

$$E\left[\left(\frac{\partial^{\iota'\boldsymbol{k}}\mu_Y(x_i, \alpha)}{\partial \alpha^{\boldsymbol{k}}}\right)^2\right] < \infty, \; E\left[\left(\frac{\partial^{\iota'\boldsymbol{k}}\sigma_Y^2(x_i, \alpha)}{\partial \alpha^{\boldsymbol{k}}}\right)^2\right] < \infty,$$

where $k = (k_1, \ldots, k_{d_\theta})$,

$$\begin{aligned} \frac{\partial^{\iota'\boldsymbol{k}}\mu_Y(x_i, \alpha)}{\partial \alpha^{\boldsymbol{k}}} &= \frac{\partial^{\iota'\boldsymbol{k}}\mu_Y(x_i, \alpha)}{\partial \alpha_1^{k_1} \ldots \partial \alpha_{d_\alpha}^{k_{d_\alpha}}}, \text{ and} \\ \frac{\partial^{\iota'\boldsymbol{k}}\sigma_Y^2(x_i, \alpha)}{\partial \alpha^{\boldsymbol{k}}} &= \frac{\partial^{\iota'\boldsymbol{k}}\sigma_Y^2(x_i, \alpha)}{\partial \alpha_1^{k_1} \ldots \partial \alpha_{d_\alpha}^{k_{d_\alpha}}}. \end{aligned}$$

Proposition 6 *Under Assumptions 1 and 2, replacing α by the restricted maximum likelihood estimator under H_0, $\hat{\alpha}$, does not alter the expressions of the score-type tests or their asymptotic distributions.*

In practice, y_i is simply replaced by $\hat{y}_i = [y_i - \mu_Y(x_i, \hat{\alpha})]/\sqrt{\sigma_Y^2(x_i, \hat{\alpha})}$ in the expressions for the different test statistics discussed in the previous section.

Proposition 6 is reminiscent of Proposition 3 in Fiorentini and Sentana (2007), who proved that when a researcher estimates a multivariate parametric location-scale model with a parametric distribution for the innovations that nests the multivariate normal, including mixtures of normals as a particular case, the (scaled, average) scores of the mean and variance parameters are asymptotically independent of the (scaled, average) scores of the shape parameters when the true distribution is in fact Gaussian. However, their proof assumes a regular model in which the usual information matrix has full rank.

5 Monte Carlo evidence

In this section, we assess the finite sample performance of our proposed tests by means of several extensive Monte Carlo exercises. The composite null hypothesis is a normal distribution with unknown mean μ and variance σ^2, while the alternative is a mixture of two normal distributions with either different means, different variances, or different means and variances. In addition, we compare our tests to the LR test, the EM tests by Chen and Li (2009) and Kasahara and Shimotsu (2015), whose asymptotic distributions is χ_2^2, and some popular nonparametric procedures based on either the empirical cumulative distribution function (cdf) or the characteristic function. Specifically, we look at the Kolmogorov-Smirnov (KS) test and the continuum of moments-test proposed in Amengual, Carrasco and Sentana (2020) (ACS).

In this context, the LR test effectively reduces to

$$LR_j = 2\left[\sup_{\vartheta\in\Theta_j'} \mathcal{L}_n(\vartheta,\hat{\mu}_n,\hat{\sigma}_n^2) - \mathcal{L}_n\left(0,0,1,\hat{\mu}_n,\hat{\sigma}_n^2\right)\right] \quad \text{with} \quad \mathcal{L}_n\left(\vartheta,\hat{\mu}_n,\hat{\sigma}_n^2\right) = \sum_{i=1}^{n} \ell_i\left(\vartheta,\hat{\mu}_n,\hat{\sigma}_n^2\right),$$

where $\ell_i\left(\vartheta,\hat{\mu}_n,\hat{\sigma}_n^2\right)$ is the log-likelihood contribution coming from the i^{th} standardized observation

$$\hat{y}_i = \frac{y_i-\hat{\mu}_n}{\hat{\sigma}_n} \quad \text{with} \quad \hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} y_i \quad \text{and} \quad \hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{\mu}_n)^2.$$

To calculate the maximizers of the unrestricted log-likelihood function, we use the GlobalSearch Toolbox in Matlab with initial value $(0,0,1/2)$, which in turn chooses another 1,000 combinations of δ, $\varkappa$ and λ. We have also tried the analogous optimization of the reparametrized log-likelihood function

$$\sup_{\theta\in\Theta_j} \sum_{i=1}^{n} l_i(\delta,\kappa,\lambda,\hat{\mu}_n,\hat{\sigma}_n^2).$$

Finally, we consider as initial values the maximizers of the eighth-order expansion of the log-likelihood function too. Specifically, for each Θ_j, we use:

- Initial value 1: $\left(\delta_n^*,\kappa_n^*-(2\lambda_n^*-1)(\delta_n^*)^2/3,\lambda_n^*\right)$, where δ_n^*, κ_n^* and λ_n^* are defined in Step 5 of the proof of Proposition 1.
- Initial value 2: $(\delta_b,\kappa_b-(2\lambda_b-1)\delta_b^2/3,\lambda_b)$, where

  $$(\delta_b,\kappa_b) \in \underset{(\delta,\kappa,1)\in\Theta_j,\delta^2+\kappa^2>10^{-3}}{\arg\max} \frac{\left(\lfloor \partial L(\delta,\kappa,1)/\partial\lambda \rfloor_-\right)^2}{V(\delta,\kappa)}$$

  and

  $$\lambda_b = \max\left\{1+\frac{1}{n}\left\lfloor \frac{1}{V(\delta_b,\kappa_b)} \frac{\partial L(\delta_b,\kappa_b,1)}{\partial\lambda}\right\rfloor_-, \frac{1}{2}\right\}.$$

It turns out, though, that maximizing the original likelihood using GlobalSearch with initial value (0,0,1/2) yielded the largest criterion function among all these possibilities

As for the other tests that we use for comparison purposes, we proceed as follows. For Kasahara and Shimotsu (2015), we fix the number of iterations K to 3, the initial value for λ (α in their notation) to 0.5, the penalty term in the penalized likelihood function as in their expression (22) with $a_n = 0.25$, and all tuning parameters to the values suggested in their paper. As for Chen and Li (2009), we use the code provided by the authors in which the number of iterations is set to $K = 2$, the initial values for λ (again, α in their notation) to $(0.1, 0.3, 0.5)$, while the penalty term is the same as in Kasahara and Shimotsu (2015), but with $a_n = 0.2 + e^{-1.410209-114.433126/n}$. In turn, we compute the KS statistic on the basis of the probability integral transforms of the standardized observations obtained through the standard normal cumulative distribution function (cdf). Finally, for the ACS test, we fix the Tikhonov regularization parameter α to 0.01 and the scale parameter ω^2 of the Gaussian density used to define distances and inner products in a suitable L^2-type Hilbert space to 1 in view of the simulation results in Amengual, Carrasco and Sentana (2020).

In all cases, we compute empirical critical values using the following parametric bootstrap procedure. First, we generate $y_1, \ldots, y_n$ as iid $N(0,1)$ and calculate the test statistics based on the observations standardized with the estimated mean and variance in that sample, restricting the parameter values over which we compute the supremum to $|\delta| \leq \bar{\delta} = 2$ and $|\varkappa| \leq \bar{\varkappa} = 1/2$. We then repeat this 10,000 times to estimate the $1-\alpha$ quantile of the distribution of our test statistics in samples of size n, which we then use as "exact" (up to Monte Carlo error) critical values. In contrast, the use of asymptotic critical values led to substantial size distortions under the null in simulations available on request.

To assess the size-corrected power of the different tests, we generate $y_1, \ldots, y_n$ from a standardized mixture of two normal distributions with several combinations of λ, δ and $\varkappa$ that include symmetric mixtures with outliers ($\varkappa > 0$), as well as asymmetric ones ($\delta \neq 0$). Then, for each sample we standardize the observations and calculate the test statistics as before, repeating this 10,000 times. Finally, we compute the corresponding rejection rates using the empirical critical values obtained by means of the parametric bootstrap procedure in the previous paragraph.

Rejection rates for sample sizes $n = 500$ (Panel A) and $n = 125$ (Panel B) are reported in Table 1. We include results for LM_j (denoted by sup in the table), $LM_{a,j}$ and $LM_{b,j}$, for $j = 1, 2, 3$ (Θ_j in the table), whenever different. Note that $LM_{a,1}$ is called JB in the table because it coincides with Jarque and Bera's test. Moreover, $LM_{b,1}$ and $LM_{b,3}$ are omitted from the table because they coincide with LM_1 and LM_3, respectively. The upper subpanels contain results for different combinations of δ and $\varkappa$ when $\lambda = 0.975$, while the lower ones do the same but when the mixing probability is 0.75, with the values δ and $\varkappa$ chosen so that the skewness and kurtosis in both subpanels coincide. As for the competitors, CL and KaSh refer to the testing procedures proposed in Chen and Li (2009) and Kasahara and Shimotsu (2015), respectively, while KS denotes the Kolmogorov-Smirnov test and ACS the CGMM test proposed in Amengual, Carrasco and Sentana (2020). As a guide, we also include two columns reporting the third and fourth moments of the alternative DGPs that we consider.

By and large, the results are very encouraging. When focusing on the parameter space Θ_1, our LM_1 test performs similarly to the usual Jarque and Bera's test. In turn, LM_2 clearly dominates both $LM_{a,2}$ and $LM_{b,2}$, while LM_3 is better than $LM_{a,3}$, as expected. In addition, the relative performance of the tests for different Θ's is in line with the alternative DGPs we consider.

Regarding the competitors proposed by Chen and Li (2009) and Kasahara and Shimotsu (2015), our proposal outperforms them for most of the Θ_i's, particularly for the alternatives with λ close to one, which confirms the importance of covering the entire parameter space. As for the consistent tests, the ACS does a decent job, beating both the LR and our score-type tests for the symmetric alternative hypothesis we consider.

We also assess the asymptotic equivalence between our LM test and the LR test by computing Gaussian rank correlation coefficients (see Amengual, Tian and Sentana (2022)), which are robust to the presence of unusually large values. Specifically, when $n = 125$ (500) we obtain .90, .88 and .86 (.93, .90 and .86) for Θ_1, Θ_2 and Θ_3, respectively.

Finally, we can confirm that computing times for the score-tests are significantly smaller than for the LR tests, taking 0.59, 0.62 and 0.27 seconds per simulation when $n = 500$ versus 1.57, 1.20 and 1.53 seconds for Θ_1, Θ_2 and Θ_3, respectively. Nevertheless, these figures underestimate the numerical advantages of our proposed tests in practice for two different reasons. First, the location-scale model that we have considered in this section only contains two parameters, unlike more realistic empirical models such as the one considered in the next section, which typically contain many more parameters that will have to be estimated under the alternative too. Second, supplemental appendix E7 of Fiorentini and Sentana (2021) shows that the ML estimators of the unconditional mean and variance parameters μ and σ^2 in a given sample numerically coincide with the sample mean and variance (with denominator n) of the observations. As a result, the criterion function maximized with respect to the shape parameters δ, κ and λ keeping μ and σ^2 fixed at those restricted ML estimators coincides with the criterion function maximized over all five parameters.

6 Empirical application to wage determinants

As is well known, the popular Mincer (1974) regression equation explains individual workers' (log) earnings of as a function of their education, measured by the number of years of schooling, and their experience, usually captured by a quadratic polynomial to reflect skill depreciation. The rationale for these variables is that labor earnings may be regarded as the returns to human capital, with education and on the job-training two different forms of investment in it.

But a simple Mincer equation fails to capture cross-sectional heterogeneity in the earnings of workers with identical schooling and experience. For example, female MBAs might earn noticeably less than male MBAs with the same number of years of experience. For that reason, empirical Mincer earnings functions often include several dummy variables, like gender or race, aimed to capture part of that heterogeneity. Formally, the gender dummy regression coefficient can be understood as the proportional decrease in labor earnings for a woman relative to a man with the same schooling and experience profile. Not surprisingly, earnings discrimination analysis often focuses precisely on the statistical significance of this regression coefficient.

But another crucial determinant of earnings is innate ability, for which data is regrettably inexistent in most labor surveys.[7] Given the dummy representation of a discrete mixture that we have exploited in our tests, a mixture model for the residuals of the Mincer equation seems very adequate to capture the possible existence of different underlying groups (or categories) of workers with noticeably different ability characteristics.[8]

Chapter 5 of Berndt (1991) contains not only a detailed analysis of the issues that arise in estimating the determinants of labor earnings, but also a random sample of 534 observations from the May 1985 issue of the Current Population Survey compiled by the US Bureau of Census. Given the illustrative nature of our analysis, we estimate by OLS the following baseline specification with all the observations in this dataset:

$$\ln w = \alpha_C + \alpha_F FE + \alpha_O OTHERS + \varepsilon,$$

where w is earnings, FE the female dummy variable, and $OTHERS$ includes dummy variables for union status, blacks, Hispanics, years of education, years of experience, its square and an interaction term between schooling and experience. In addition, we estimate the same regression specification using exclusively female and male subsamples separately after dropping FE to avoid collinearity. For each of those three empirical specifications, we test whether the residual follows a normal distribution with 0 mean and unknown variance σ^2.

Unfortunately, we cannot use the parametric bootstrap to compute the critical values as we did in our Monte Carlo simulations because of the presence of regressors. For that reason, we use the following semiparametric bootstrap procedure:

1. Regress Y $(= \ln w)$ on the explanatory variables (X) and obtain the ordinary least-squares estimates $\hat{\alpha}$, $\hat{\sigma}^2$, and the OLS residual $\hat{\varepsilon}$.

2. Calculate the test statistic (denoted $\widehat{T}$ for simplicity) using $\hat{\varepsilon}$.

3. Using random sampling with replacement to nonparametrically bootstrap the regressors, X_b, and then construct $Y_b = X_b\hat{\alpha} + \hat{\sigma}\varepsilon_b$, where $\varepsilon_b|(Y, X) \sim iid\ N(0, 1)$.

4. Regress Y_b on X_b and get $\hat{\alpha}_b$ and $\hat{\varepsilon}_b$.

5. Calculate the test statistic T_b with input $\hat{\varepsilon}_b$.

[7] Griliches and Mason (1972) constitute an important exception, as they had data on both earnings and IQ scores for the individuals in their sample. Somewhat surprisingly, though, they found that their ability measures were essentially uncorrelated with schooling, which means that the omitted variable bias in measuring the returns to education was negligible.

[8] See Bonhomme and Manresa (2015) for a closely related approach in panel data.

6. Repeat 10,000 times steps 2 to 5 and compute the bootstrap p-value as $\frac{1}{B}\sum_{b=1}^{B}\mathbf{1}[T_b > \hat{T}]$.

The results of the empirical application are displayed in Table 3. The first column includes results for the full sample, and the second and third ones for men and women separately. On the basis of the p-values, we can see that the distribution of wages for the entire sample, conditional on the regressors, is leptokurtic but apparently symmetric. However, when we distinguish between males and females, some asymmetry appears, with positive skewness for men and negative skewness for women. Moreover, our tests reject the null hypothesis of normality against the normal mixture, which suggests that some unobserved heterogeneity remains in both samples.

7 Conclusions and directions for further research

This paper presents score-type tests for normality against normal mixtures with different means or variances. Our tests, which are robust to the sampling uncertainty resulting from the estimation of the conditional mean and variance parameters used to construct standardized residuals, are asymptotically equivalent to the LR test.

For illustrative purposes, we focus on mixtures of two normal distributions. Considering more than two categories would represent an interesting extension. We could also explore procedures to determine the number of components in normal mixture models, as in Kasahara and Shimotsu (2015). We have restricted ourselves to serially independent observations, but the underlying regimes may be somewhat persistent in many macroeconomic and financial applications. An extension of our work to the Markov-switching models studied by Carrasco, Hu and Ploberger (2014) and Qu and Fan (2021) provides another promising route for future research.

It would also be interesting to consider other distributions besides the normal. In fact, the normal distribution is very special and some of the difficulties we have dealt with, such as the singularity of the information matrix, may not arise with other mixtures. On the other hand, scale mixtures of univariate normals give rise to mixtures of chi-square distributions with 1 degree of freedom for the squares, and the same happens in the multivariate case if we consider the exponents of the multivariate normal density, except that the degrees of freedom of the chi-squares will coincide with the dimension of the random vectors. Therefore, it should be possible to test for mixtures of two chi-squares using our existing results. We are currently exploring some of these interesting research avenues.

Appendix: Proofs

The proofs of our main theorems use some lemmas which we present and prove in the Supplemental Appendix D. We will also make extensive use of the following notation:

1. the stochastic sequence a_n is "bounded in probability", or $O_p(1)$, when $\forall \epsilon > 0$, there exists M such that $\Pr(|a_n| < M) \geq 1 - \epsilon$ for all n;

2. the sequence of events A_n holds "infinitely often" (i.o.) when the cardinality of the set $\{n : A_n \text{ holds}\}$ is infinite; and

3. A_n holds ultimately (all but finite) when there exists N such that $\{n : A_n \text{ holds}\} = \{n : n \geq N\}$, with $N < \infty$. "ultimately" is denoted "ult." in the sequel.

Overview of the proof of Proposition 1

We find the score-type test statistic that is asymptotically equivalent to

$$2\left[\sup_{\theta \in \mathcal{P}_a} L_n(\theta) - L_n(0,0,1)\right],$$

where $\mathcal{P}_a$ satisfies that $(0,0,1) \in \mathcal{P}_a \subseteq \Theta_1$. Notice that in the proof, we use $\mathcal{P}_a$ as the parameter space, but we could, when required, change from $\mathcal{P}_a$ to $\mathcal{P}_{a,j}$ for $j = 1,2,3$. With a slight abuse of notation, we also define

$$\begin{aligned} LR_n(\theta) &= 2\left[L_n(\theta) - L_n(0,0,\lambda)\right] \quad \text{and} \\ LM_n^a(\theta) &= 2\frac{H_{3,n}}{\sqrt{n}} w_1 - V_3 w_1^2 + 2\frac{H_{4,n}}{\sqrt{n}} w_2 - V_4 w_2^2, \end{aligned} \tag{15}$$

where

$$w_1 = -\frac{\lambda}{2}\sqrt{n}(1-\lambda)\delta\kappa \quad \text{and} \quad w_2 = -\frac{\lambda(1-\lambda+\lambda^2)}{36}\sqrt{n}(1-\lambda)\delta^4 + \frac{\lambda}{8}\sqrt{n}(1-\lambda)\kappa^2.$$

Moreover, note that $L_n(\delta,\kappa,1) = L_n(0,0,\lambda)$.

There are five steps in the proof:

1. For all sequences of $\theta_n \in \Theta$ with $(\delta_n, \kappa_n) \xrightarrow{p} 0$, we have that

$$LR_n(\theta_n) = LM_n^a(\theta_n) + o_p[h_n(\theta_n)],$$

   where $h_n(\theta_n) = \max\left\{1, n(1-\lambda_n)^2\delta_n^8, n(1-\lambda_n)^2\delta_n^2\kappa_n^2, n(1-\lambda_n)^2\kappa_n^4\right\}$.

2. Defining $\theta_n^{LM} = (\delta_n^{LM}, \kappa_n^{LM}, \lambda_n^{LM}) \in \text{argmax}_{\theta\in\Theta} LM_n^a(\theta)$, we show that $(\delta_n^{LM}, \kappa_n^{LM}) \xrightarrow{p} 0$ and $h_n(\theta_n^{LM}) = O_p(1)$.

3. Defining $\theta_n^{LR} = (\delta_n^{LR}, \kappa_n^{LR}, \lambda_n^{LR}) \in \text{argmax}_{\theta\in\Theta} LR_n(\theta)$, we also show that $(\delta_n^{LR}, \kappa_n^{LR}) \xrightarrow{p} 0$ and $h_n(\theta_n^{LR}) = O_p(1)$.

4. We then prove that $LR_n(\theta_n^{LR}) = LM_n^a(\theta_n^{LM}) + o_p(1)$.

5. We finally simplify $LM_n^a(\theta_n^{LM})$ to LM_{a1} (respectively, LM_{a2} and LM_{a3}) in $\mathcal{P}_{a1}$ (respectively, $\mathcal{P}_{a2}$ and $\mathcal{P}_{a3}$).

The detailed steps can be found in Supplemental Appendix A. □

Proof of Proposition 2

Regarding part (1), by Theorem 10.2 of Pollard (1990) (see also Andrews (2001)),

$$\frac{1}{\sqrt{n}}\sum_i \frac{\partial l_i}{\partial \lambda}(.,1) \Rightarrow G(.)$$

if (i) B (the set within which the index lies) is totally bounded, (ii) the finite dimensional distributions of $\frac{1}{\sqrt{n}}\sum_i \frac{\partial l_i}{\partial \lambda}(.,1)$ converge to those of $G(.)$, (iii) $\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\partial l_i}{\partial \lambda}(.,1) : n \geq 1\right\}$ is stochastically equicontinuous.

(i) is satisfied because $\beta \equiv (\delta,\kappa) \in B = \left\{(\delta,\kappa) : (\delta,\kappa,1) \in \mathcal{P}_{b,1} \text{ and } \sqrt{\delta^2+\kappa^2} \geq \epsilon\right\}$ and B is compact.

(ii) The process $\frac{\partial l_i}{\partial \lambda}(.,1)$ is *iid* with mean 0. Moreover,

$$E\sup_{\beta\in B}\left|\frac{\partial l}{\partial \lambda}(\beta,1)\right| < \infty. \tag{16}$$

Indeed, the absolute value of the score involves a constant, a linear combination of $|y_i|$ and y_i^2, and finally an exponential term. By the definition of B, we cannot have $\delta = 0$ and $\kappa = 0$ simultaneously. Below, we use the notation y for y_i while $\varkappa$ denotes $\kappa - \delta^2/3$. As κ and δ belong to compact sets, so does $\varkappa$. Hence, we can write $\varkappa \in [-\overline{\varkappa},\overline{\varkappa}]$. Moreover, $1-e^{-\varkappa} \leq 1-e^{-\overline{\varkappa}} < 1$ and

$$\begin{aligned}
\frac{1}{\sqrt{e^{\varkappa}}}\exp\left[\frac{1}{2}\left\{y^2-\frac{[y+\delta]^2}{e^{\varkappa}}\right\}\right] &= e^{-\varkappa/2}\exp\left[-\frac{1}{2}\frac{(1-e^{\varkappa})}{e^{\varkappa}}y^2\right]\exp\left(-\frac{y\delta}{e^{\varkappa}}\right)\exp\left(-\frac{\delta^2}{2e^{\varkappa}}\right)\\
&= e^{-\varkappa/2}\exp\left[\frac{1}{2}\left(1-e^{-\varkappa}\right)y^2\right]\exp\left(-\frac{y\delta}{e^{\varkappa}}\right)\exp\left(-\frac{\delta^2}{2e^{\varkappa}}\right)\\
&\leq \exp\left[\frac{1}{2}\left(1-e^{-\overline{\varkappa}}\right)y^2\right]\exp\left(\frac{|y|\,|\delta|}{e^{\varkappa}}\right)\\
&\leq \exp\left[\frac{1}{2}\left(1-e^{-\overline{\varkappa}}\right)y^2\right]\exp\left(|y|\,e^{\overline{\varkappa}}\,|\delta|\right)\\
&\equiv g^*(y).
\end{aligned} \tag{17}$$

Note that $E[g^*(y)]$ is finite because $1-e^{-\overline{\varkappa}} < 1$. So we can major $\left|\frac{\partial l_i}{\partial \lambda}(\beta,1)\right|$ by terms which do not depend on β and have finite expectations.

By (16), the martingale difference central limit theorem of Billingsley (1968, Theorem 3.1) implies that each of the finite dimensional distributions of $\frac{1}{\sqrt{n}}\sum_i \frac{\partial l_i}{\partial \lambda}(.,1)$ converges in distribution to a multivariate normal distribution whose covariance matrix is characterized by (8).

(iii) Let $\nu_n(\beta) = \frac{1}{\sqrt{n}}\sum_i \frac{\partial l_i}{\partial \lambda}(\beta, 1)$. A process $\nu_n(\beta)$ is stochastically equicontinuous if for all $\varepsilon > 0$, there exists $c > 0$ such that

$$\overline{lim}_{n\to\infty} P\left[\sup_{\|\beta_1-\beta_2\|\leq c} |\nu_n(\beta_1) - \nu_n(\beta_2)| > \varepsilon\right] < \varepsilon.$$

To establish that the process $\nu_n(\beta)$ is stochastically equicontinuous, we use Theorem 1 of Andrews (1994). First, we use the notation f for $\nu_n(\beta) = \frac{1}{\sqrt{n}}\sum_i f(y_i, \beta)$ and show that f belongs to the type II class of functions defined in Andrews (1994, p.2270). This is the class of Lipschitz functions in β, which is such that

$$|f(.,\beta_1) - f(.,\beta_2)| \leq M(.)\|\beta_1 - \beta_2\|, \text{ for all } \beta_1, \beta_2 \in B.$$

But

$$\begin{aligned} f(y,\beta_1) - f(y,\beta_2) &= \frac{e^{\varkappa_2} - e^{\varkappa_1}}{2} - e^{-\varkappa_1/2}\exp\left\{\frac{1}{2}[y^2 - (y+\delta_1)^2 e^{-\varkappa_1}]\right\} \\ &+ e^{-\varkappa_2/2}\exp\left\{\frac{1}{2}[y^2 - (y+\delta_2)^2 e^{-\varkappa_2}]\right\} + (\delta_2 - \delta_1)y \\ &+ (e^{\varkappa_1} - e^{\varkappa_2})\frac{y^2}{2} + \frac{(\delta_1^2 - \delta_2^2)}{2}(y^2 - 1). \end{aligned}$$

Using the mean-value theorem, we have

$$e^{\varkappa_2} - e^{\varkappa_1} = e^{\tilde{\varkappa}}(\varkappa_2 - \varkappa_1),$$

where $\tilde{\varkappa}$ lies between $\varkappa_1$ and $\varkappa_2$. Hence, $|e^{\varkappa_2} - e^{\varkappa_1}| = e^{\tilde{\varkappa}}|\varkappa_2 - \varkappa_1| \leq e^{\overline{\varkappa}}|\varkappa_2 - \varkappa_1|$. Let

$$g(y,\beta) = -e^{-\varkappa/2}\exp\left\{\frac{1}{2}[y^2 - (y+\delta)^2 e^{-\varkappa}]\right\}.$$

The mean-value theorem gives

$$\begin{aligned} g(y,\beta_1) - g(y,\beta_2) &= \frac{1}{2}\left[e^{-\tilde{\varkappa}}(y+\tilde{\delta})^2 - 1\right] g(y,\widetilde{\beta})(\varkappa_1 - \varkappa_2) - (y+\tilde{\delta})e^{-\tilde{\varkappa}} g(y,\widetilde{\beta})(\delta_1 - \delta_2) \\ |g(y,\beta_1) - g(y,\beta_2)| &\leq \frac{1}{2}[e^{\tilde{\varkappa}}(y^2 + 2|y||\overline{\delta}| + |\overline{\delta}|^2) + 1]g^*(y)|\varkappa_1 - \varkappa_2| \\ &+ (|y| + |\overline{\delta}|)e^{\tilde{\varkappa}} g^*(y)|\delta_1 - \delta_2|, \end{aligned}$$

where $\widetilde{\beta} = (\tilde{\delta}, \tilde{\kappa})$, $\tilde{\delta}$ is between δ_1 and δ_2, and g^* is defined in (17). Note that $|\delta_1 - \delta_2| \leq \|\beta_1 - \beta_2\|$ and $|\varkappa_1 - \varkappa_2| \leq \|\beta_1 - \beta_2\|$. Hence, f is Lipschitz with $M(y) = c_0 + c_1 y + c_2 y^2 + c_3|y| g^*(y) + c_4 y^2 g^*(y)$ for some constants c_0, c_1, c_2, c_3 and c_4. Now, to apply Theorem 1 of Andrews (1994), we need to check his Assumptions A, B, and C. Specifically, Assumption A: the class of functions f satisfies Pollard's entropy condition with some envelope $\bar{M}$. This is satisfied with $\bar{M} = 1 \vee \sup|f| \vee M(.)$ by Theorem 2 of Andrews (1994) because f is Lipschitz. Similarly, Assumption B:

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \sum_{i}^{n} E\bar{M}^{2+v}(y_i) < \infty \text{ for some } v > 0.$$

This condition is also satisfied because y_i is a standard normal random variable (r.v.). In turn, Assumption C: $\{y_i\}$ is an m-dependent triangular array of r.v's holds because $\{y_i\}$ is *iid*. Finally, stochastic equicontinuity follows from Theorem 1 in Andrews (1994).

As for part (2) of the proposition, expressions (a) and (b) are direct consequences of part (1) and the continuous mapping theorem. In turn, expression (c) follows from Andrews (2001). To see this, we need to check the assumptions in Andrews (2001), whose notation is such that θ is our λ and π is our (δ, κ). Let l_i denote the log-likelihood of y_i. Note that $\lambda + (1-\lambda)\exp(\varkappa) \leq 1 + \exp(\overline{\varkappa})$ and $1 + \lambda(1-\lambda)\delta^2 \leq 1 + \delta^2/4 \leq 1 + \overline{\delta}^2$. As a consequence, $\sigma_1^* \geq [(1+\overline{\delta}^2)(1+\exp(\overline{\varkappa}))]^{-1} > 0$ and $\sigma_2^* \geq \exp(-\overline{\varkappa})[(1+\overline{\delta}^2)(1+\exp(\overline{\varkappa}))]^{-1} > 0$.

To verify Assumption 1*(a), it suffices to apply the uniform law of large numbers (see Lemma 2.4 of Newey and McFadden (1994)), which holds because $\{l_i\}$ is *iid*, continuous in both λ and $\beta \equiv (\delta, \kappa)$ with probability one, and

$$E \sup_{\lambda\in[0,1],\beta\in B} |l_i(\beta,\lambda)| \leq \sup_{\lambda\in[0,1],\beta\in B} \ln\left\{\frac{1}{\sqrt{2\pi}\sigma_1^*} + \frac{1}{\sqrt{2\pi}\sigma_2^*}\right\} < \infty.$$

Moreover, the limit $\sum_i l_i(\beta,\lambda)/n$ is $E[l_i(\beta,\lambda)] \equiv l(\beta,\lambda)$, which does not depend on β when $\lambda = 1$.

To verify Assumption 1*(b), we need to show that $l(\beta,\lambda)$ is maximized over $[0,1]$ at $\lambda_0 = 1$ for each $\beta \in B$. By the properties of maximum likelihood estimators (see Theorem 2.5 of Newey and McFadden (1994)), it suffices to check that $P[l_i(\beta,\lambda) \neq l_i(\beta_0,\lambda_0)] > 0$ for any $\beta \neq \beta_0$ and $\lambda \neq \lambda_0 = 1$, which is true here.

Assumption 2^{2^*}(a) is clearly satisfied for $\Theta^+ = (1-\varepsilon, 1)$.

As for Assumption 2^{2^*}(b), it is easy to check that $l_i(\beta,\lambda)$ has left and right partial derivatives with respect to λ on Θ^+, $\forall\beta \in B$.

Regarding Assumption 2^{2^*}(c), we can show that for all $\gamma_n \to 0$,

$$\sup_{\lambda\in[0,1]:\|\lambda-1\|\leq\gamma_n} \left\| \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial^2}{\partial\lambda^2}l_i(\beta,\lambda) - \frac{\partial^2}{\partial\lambda^2}l_i(\beta,1)\right]\right\| = o_{p\beta}(1)$$

where $X_{n\beta} = o_{p\beta}(1)$, implies that $\sup_{\beta\in B}\|X_{n\beta}\| = o_p(1)$. This condition is tedious to check but does not raise any special difficulty, so the details are omitted.

Assumption 3^* holds by of part (A) of Proposition 2. Assumption 5 is satisfied for $B_n =$

$b_n = \sqrt{n}$ and $\Lambda = \mathbb{R}^-$. Assumption 6 holds because $\mathbb{R}^-$ is convex.

Assumptions 7 and 8 hold with $\Lambda_\beta = \mathbb{R}^-$ and with the fact that β in Andrews (2001)'s notation corresponds to our λ, while his (δ, ψ) are absent in our setting.

Assumptions 9 and 10 are satisfied. Assumptions 1o and 4o hold trivially because the restricted estimator is $\lambda = 1$ and therefore not random.

By Theorem 4 and the remark at the bottom of p. 719 of Andrews (2001), it follows that $LR_{b,1}^B = LM_{b,1}^B + o_p(1)$. □

Overview of the proof of Proposition 3

Part (a) follows from the results of Proposition 2 and the continuous mapping theorem.

In part (b), we look for the score-type test statistic that is asymptotically equivalent to

$$2\left[\sup_{\theta\in\mathcal{P}_b} L_n(\theta) - L_n(0,0,1)\right]$$

where $\mathcal{P}_b$ satisfies that $(0,0,1) \in \mathcal{P}_b \subseteq \Theta_1$. Notice that in the following proof we use $\mathcal{P}_b$ as the parameter space, but we could, if necessary, replace $\mathcal{P}_b$ with $\mathcal{P}_{b,k}$ for $k = 1,2,3$. For $\theta \in \mathcal{P}_b$, define

$$LR_n(\theta) = 2\left[L_n(\theta) - L_n(0,0,1)\right]$$

and for $\theta \in \mathcal{P}_b \backslash \{(0,0,1)\}$, let

$$\begin{aligned} LM_n^b(\theta) &= \frac{2}{\sqrt{n}} \frac{\partial L_n(\delta,\kappa,1)}{\partial \lambda} \sqrt{n}(\lambda - 1) - V(\delta,\kappa) n(\lambda-1)^2 \quad \text{and} \\ V_b(\delta,\kappa) &= E\left[\left(\frac{\partial l(\delta,\kappa,1)}{\partial\lambda}\right)^2\right]. \end{aligned}$$

We will show that the LR test statistic is asymptotically equivalent to the following score-type statistic:

$$\sup_{\theta\in\mathcal{P}_b} LR_n(\theta) = \frac{1}{n} \sup_{\delta,\kappa:(\delta,\kappa,1)\in\mathcal{P}_b\backslash(0,0,1)} \frac{\left(\min\left\{\partial L_n(\delta,\kappa,1)/\partial\lambda, 0\right\}\right)^2}{V(\delta,\kappa)} + o_p(1).$$

The LM statistic is usually constructed based on the first two terms of the Taylor expansion. A third-order Taylor expansion of $l(\theta)$ gives

$$l(\delta,\kappa,\lambda) - l(\delta,\kappa,1) = \frac{\partial l(\delta,\kappa,1)}{\partial\lambda}(\lambda-1) + \frac{1}{2}\frac{\partial^2 l(\delta,\kappa,1)}{\partial\lambda^2}(\lambda-1)^2 + \frac{1}{3!}\frac{\partial^3 l(\delta,\kappa,\tilde{\lambda})}{\partial\lambda^3}(\lambda-1)^3.$$

It is then easy to verify that $\partial l(\delta,\kappa,1)/\partial\lambda = 0$ at $(\delta,\kappa) = 0$, which confirms the singular information matrix problem. Moreover, the limit

$$\lim_{(\delta,\kappa)\to 0} \frac{1}{\sqrt{V(\delta,\kappa)}} \frac{\partial l(\delta,\kappa,1)}{\partial\lambda}$$

does not exist because its value depends on the direction of (δ,κ) (see Supplemental Appendix B for an example). One way to circumvent this problem is to normalize $\partial l(\delta,\kappa,1)/\partial\lambda$ by a function of (δ,κ) and further reparametrize the model. To be more specific, for $\delta^2+\kappa^2>0$, let

$$\tau = \max\left\{\left|\frac{1}{36}\delta^4-\frac{1}{8}\kappa^2\right|,\left|\frac{1}{2}\delta\kappa\right|\right\}, \tag{18}$$

$$\varphi = \frac{\min\left\{\left|\frac{1}{36}\delta^4-\frac{1}{8}\kappa^2\right|,\left|\frac{1}{2}\delta\kappa\right|\right\}}{\max\left\{\left|\frac{1}{36}\delta^4-\frac{1}{8}\kappa^2\right|,\left|\frac{1}{2}\delta\kappa\right|\right\}} \text{ and} \tag{19}$$

$$\eta = \max\left\{\left|\frac{1}{36}\delta^4-\frac{1}{8}\kappa^2\right|,\left|\frac{1}{2}\delta\kappa\right|\right\}(1-\lambda). \tag{20}$$

Note that $\tau>0$ if and only if $\delta^2+\kappa^2>0$. Additionally, we can normalize the score by τ as follows: if $\left|\frac{1}{36}\delta^4-\frac{1}{8}\kappa^2\right|\geq\left|\frac{1}{2}\delta\kappa\right|$,

$$\lim_{\tau\to 0}\tau^{-1}\frac{\partial l(\delta,\kappa,1)}{\partial\lambda}=\text{sign}\left(\frac{1}{36}\delta^4-\frac{1}{8}\kappa^2\right)h_4+\text{sign}\left(\frac{1}{2}\delta\kappa\right)h_3\varphi,$$

and if $\left|\frac{1}{36}\delta^4-\frac{1}{8}\kappa^2\right|\leq\left|\frac{1}{2}\delta\kappa\right|$,

$$\lim_{\tau\to 0}\tau^{-1}\frac{\partial l(\delta,\kappa,1)}{\partial\lambda}=\text{sign}\left(\frac{1}{36}\delta^4-\frac{1}{8}\kappa^2\right)h_4\varphi+\text{sign}\left(\frac{1}{2}\delta\kappa\right)h_3.$$

To further simplify the notation, we also reparametrize from θ to $d=(\eta,\tau,\varphi)$.

To guarantee that there is a one to one mapping from θ to d, we further partition the parameter space into the following sets. Let

$$\begin{aligned}
A_{10} &= \left\{(\delta,\kappa,\lambda)\in\mathcal{P}_b:\left|\frac{1}{2}\delta\kappa\right|\leq\left|\frac{1}{36}\delta^4-\frac{1}{8}\kappa^2\right|,\delta^2+\kappa^2>0\right\},\\
A_{20} &= \left\{(\delta,\kappa,\lambda)\in\mathcal{P}_b:\frac{1}{36}\delta^4-\frac{1}{8}\kappa^2\geq 0,\delta^2+\kappa^2>0\right\},\\
A_{30} &= \left\{(\delta,\kappa,\lambda)\in\mathcal{P}_b:\kappa\geq 0,\delta^2+\kappa^2>0\right\} \quad\text{and}\\
A_{40} &= \left\{(\delta,\kappa,\lambda)\in\mathcal{P}_b:\delta\geq 0,\delta^2+\kappa^2>0\right\},
\end{aligned}$$

Define $A_{i1}=\mathcal{P}_b\backslash\left(A_{i0}\cup\{(0,0,1)\}\right)$ and let

$$\left\{A^1,\ldots,A^{16}\right\}=\left\{\cap_{i=1}^4 A_{ij_i}:(j_1,\ldots,j_4)\in\{0,1\}^4\right\}.$$

It is easy to see that

$$\sup_{\theta\in\mathcal{P}_b}LR_n(\theta)=\max_{k\leq 16}\sup_{\theta\in A^k}LR_n(\theta) \quad\text{and}\quad \sup_{\theta\in\mathcal{P}_b}LM_n^b(\theta)=\max_{k\leq 16}\sup_{\theta\in A^k}LM_n^b(\theta).$$

As a consequence, it suffices to consider the asymptotic equivalence between $\sup_{\theta\in A^k}LR_n(\theta)$ and $\sup_{\theta\in A^k}LM_n(\theta)$ for each A^k. Let

$$D^k=\left\{d=(\eta,\tau,\varphi):\text{there exists }\theta\in A^k\text{ such that (20)-(19) holds}\right\}.$$

By Lemma 4, there is a one-to-one mapping between $\theta \in A^k$ and $d \in D^k$.

Similarly, let

$$A^k_{\delta\kappa} = \{(\delta, \kappa) : \text{there exists } \lambda \text{ such that } (\delta, \kappa, \lambda) \in A^k\} \text{ and}$$
$$D^k_{\tau\varphi} = \{(\tau, \varphi) : \text{there exists } \eta \text{ such that } (\eta, \tau, \varphi) \in D^k\}.$$

We will show below the asymptotic equivalence of $\sup_{\theta \in A^1} LR_n(\theta)$ and $\sup_{\theta \in A^1} LM_n(\theta)$ for $A^1 = \cap_{i=1}^4 A_{i0}$. The proofs for the remaining 15 sets are very similar, so we omit them in the interest of space. With a slight abuse of notation, let $\delta(\tau, \varphi), \kappa(\tau, \varphi), \lambda(\eta, \tau, \varphi)$ denote the value of δ, κ, λ for given (η, τ, φ), and let $\eta(\delta, \kappa, \lambda), \tau(\delta, \kappa), \varphi(\delta, \kappa)$ denote the value of (η, τ, φ) for given (δ, κ).

For $(\tau, \varphi) \in D^1_{\tau\varphi}$, let

$$\mathcal{G}^d_n(\tau, \varphi) = \frac{1}{\sqrt{n}} \tau^{-1} \frac{\partial L_n(\delta(\tau, \varphi), \kappa(\tau, \varphi), 1)}{\partial \lambda},$$

so that

$$\lim_{\tau \to 0} \mathcal{G}^d_n(\tau, \varphi) = \frac{1}{\sqrt{n}}(H_4 + \varphi H_3).$$

Finally, let

$$LM^d_n(\eta, \tau, \varphi) = 2\mathcal{G}^d_n(\tau, \varphi)\sqrt{n}\eta - V(\tau, \varphi)n\eta^2 \text{ and}$$
$$LR^d_n(\eta, \tau, \varphi) = LR_n[\delta(\tau, \varphi), \kappa(\tau, \varphi), \lambda(\eta, \tau, \varphi)],$$

There will be four steps in the proof:

1. For all sequences of $(\eta_n, \tau_n, \varphi_n) \in D^1$ and $\eta_n \xrightarrow{p} 0$, we have that

$$LR^d_n(\eta_n, \tau_n, \varphi_n) - LM^d_n(\eta_n, \tau_n, \varphi_n) = o_p(n\eta_n^2).$$

2. Weak convergence of the process $\{\mathcal{G}^d_n(\tau, \varphi) : (\tau, \varphi) \in D^1_{\tau\varphi}\}$.

3. We prove that

$$\sup_{d \in D^1} LR^d_n(d) = \sup_{d \in D^1} LM^d_n(d) + o_p(1) = \sup_{(\tau,\varphi) \in D^1_{\tau\varphi}} \frac{\left(\min\left\{\mathcal{G}^d_n(\tau, \varphi), 0\right\}\right)^2}{V(\delta, \kappa)} + o_p(1).$$

4. Main theorem: We combine results for the 16 sets and go back to the $(\delta, \varkappa)$ space

$$\sup_{\vartheta \in \Theta'} 2\left(\mathcal{L}_n(\vartheta) - \mathcal{L}_n(0, 0, 1)\right) = \frac{1}{n} \sup_{\vartheta \in \Theta'} \frac{\left(\min\left\{\partial \mathcal{L}_n(\delta, \varkappa, 1)/\partial \lambda, 0\right\}\right)^2}{V(\delta, \varkappa)} + o_p(1).$$

The detailed steps can be found in Supplemental Appendix B. □

Proof of Proposition 4

To show (12) for $j=1$, note that for $k_1 \in \mathbb{R}$,

$$\lim_{\varepsilon\to 0}\left\lfloor \frac{\mathcal{G}_n(\varepsilon,k_1\varepsilon)}{\sqrt{V(\varepsilon,k_1\varepsilon)}}\right\rfloor_-^2 = \frac{1}{n}\frac{\lfloor 4k_1H_{3,n}-k_1^2H_{4,n}\rfloor_-^2}{16k_1^2V_3+k_1^4V_4}. \tag{21}$$

In addition, let $k_1 = -4\frac{H_{4,n}V_3}{H_{3,n}V_4}$, which is well defined with probability one. Then, we can write

$$\begin{aligned}(21) &= \frac{1}{n}\frac{\lfloor 4k_1H_{3,n}-k_1^2H_{4,n}\rfloor_-^2}{16k_1^2V_3+k_1^4V_4} = \frac{1}{n}\frac{\left\lfloor -16\frac{H_{4,n}V_3^2}{H_{3,n}^2V_4}\left(\frac{H_{3,n}^2}{V_3}+\frac{H_{4,n}^2}{V_4}\right)\right\rfloor_-^2}{16^2\frac{H_{4,n}^2V_3^4}{H_{3,n}^4V_4^2}\left(\frac{H_{3,n}^2}{V_3}+\frac{H_{4,n}^2}{V_4}\right)}\\ &= \frac{1}{n}\left(\frac{H_{3,n}^2}{V_3}+\frac{H_{4,n}^2}{V_4}\right)\mathbf{1}\left[H_{4,n}\geq 0\right].\end{aligned}$$

On the other hand, for $k_2 \in \mathbb{R}$,

$$\lim_{\varepsilon\to 0}\left\lfloor \frac{\mathcal{G}_n(\varepsilon,\frac{k_2}{18}\varepsilon^3)}{\sqrt{V(\varepsilon,\frac{k_2}{18}\varepsilon^3)}}\right\rfloor_-^2 = \frac{1}{n}\frac{\lfloor H_{4,n}+k_2H_{3,n}\rfloor_-^2}{V_4+k_2^2V_3}. \tag{22}$$

Letting $k_2 = \frac{H_{3,n}V_4}{H_{4,n}V_3}$, we can write

$$\begin{aligned}(22) &= \frac{1}{n}\frac{\lfloor H_{4,n}+k_2H_{3,n}\rfloor_-^2}{V_4+k_2^2V_3} = \frac{1}{n}\frac{\left\lfloor H_{4,n}+\frac{H_{3,n}V_4}{H_{4,n}V_3}H_{3,n}\right\rfloor_-^2}{V_4+\frac{H_{3,n}^2V_4^2}{H_{4,n}^2V_3^2}V_3} = \frac{1}{n}\frac{\left\lfloor \frac{V_4}{H_{4,n}}\left(\frac{H_{4,n}^2}{V_4}+\frac{H_{3,n}^2}{V_3}\right)\right\rfloor_-^2}{\frac{V_4^2}{H_{4,n}^2}\left(\frac{H_{4,n}^2}{V_4}+\frac{H_{3,n}^2}{V_3}\right)}\\ &= \frac{1}{n}\left(\frac{H_{3,n}^2}{V_3}+\frac{H_{4,n}^2}{V_4}\right)\mathbf{1}\left[H_{4,n}< 0\right].\end{aligned}$$

Then, it is easy to show that with probability 1,

$$\begin{aligned}\sup_{|\delta|\leq\bar{\delta},|\kappa-\delta^2/3|\leq\bar{\varkappa},|\kappa|,|\delta|>0}\left\lfloor\frac{\mathcal{G}_n(\delta,\varkappa)}{\sqrt{V(\delta,\varkappa)}}\right\rfloor_-^2 &\geq \max\left\{\lim_{\varepsilon\to 0}\left\lfloor \frac{\mathcal{G}_n(\varepsilon,k_1\varepsilon)}{\sqrt{V(\varepsilon,k_1\varepsilon)}}\right\rfloor_-^2, \lim_{\varepsilon\to 0}\left\lfloor \frac{\mathcal{G}_n(\varepsilon,\frac{k_2}{18}\varepsilon^3)}{\sqrt{V(\varepsilon,\frac{k_2}{18}\varepsilon^3)}}\right\rfloor_-^2\right\}\\ &= \frac{1}{n}\left(\frac{H_{3,n}^2}{V_3}+\frac{H_{4,n}^2}{V_4}\right).\end{aligned}$$

In turn, to show (12) for $j=3$, note that

$$\begin{aligned}\mathcal{G}_n(0,\kappa) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{2}\left[(e^\kappa-1)(y_i^2-1)+2-2e^{\frac{1}{2}\left((1-e^{-\kappa})y_i^2-\kappa\right)}\right] \quad \text{and}\\ V(0,\kappa) &= \frac{1}{2}\left(-\frac{2\sqrt{2e^{-\kappa}-1}}{e^\kappa-2}+2e^\kappa-e^{2\kappa}-3\right)\end{aligned}$$

with

$$\lim_{\kappa\to 0}\frac{\mathcal{G}_n(0,\kappa)}{\sqrt{V(0,\kappa)}}=\frac{-H_{4,n}}{\sqrt{n}\sqrt{V_4}}.$$

As a consequence,

$$\sup_{|\kappa|\le\bar{\varkappa},|\kappa|>0}\left\lfloor\frac{\mathcal{G}_n(0,\kappa)}{\sqrt{V(0,\kappa)}}\right\rfloor_-^2\ge\left\lfloor\frac{-H_{4,n}}{\sqrt{n}\sqrt{V_4}}\right\rfloor_-^2=\frac{H_{4,n}^2}{nV_4}\mathbf{1}\left[H_{4,n}>0\right],$$

as desired. Results (13) and (14) follow from Propositions 2 and 3. □

Proof of Proposition 5

We show the results for $\mathcal{H}_{2n}$ first.

Contiguity. By Le Cam's first lemma (see Lemma 6.4 of van der Vaart (1998)), contiguity holds if $dP_{\beta,\lambda_n}/dP_0\xrightarrow{d}U$ under P_0 with $E(U)=1$. Let $L_n(\beta,\lambda)$ denote the joint likelihood of $y_1,\ldots,y_n$ for a given β and λ. By the mean value theorem, we have

$$L_n(\beta,\lambda)=L_n(\beta,\lambda_0)+\frac{\partial L_n(\beta,\lambda_0)}{\partial\lambda}(\lambda_n-\lambda_0)+\frac{1}{2}\frac{\partial^2 L_n(\beta,\tilde{\lambda})}{\partial\lambda^2}(\lambda_n-\lambda_0)^2,$$

where $\tilde{\lambda}$ is between λ_0 and λ_n. Replacing λ_0 by 1 and using Andrews (2001), we have

$$\begin{aligned}L_n(\beta,\lambda_n)&=L_n(\beta,1)-\frac{\partial L_n(\beta,1)}{\partial\lambda}\frac{\rho}{\sqrt{n}}-\frac{1}{2}\frac{\partial^2 L_n(\beta,\tilde{\lambda})}{\partial\lambda^2}\frac{\rho^2}{n}\\&=L_n(\beta,1)-\frac{1}{\sqrt{n}}\frac{\partial L_n(\beta,1)}{\partial\lambda}\rho-\frac{1}{2}var[G(\beta)]\rho^2+o_{p\beta}(1).\end{aligned}$$

Therefore, under H_0,

$$\begin{aligned}\frac{dP_{\beta,\lambda_n}}{dP_0}&=\exp\left\{-\frac{1}{\sqrt{n}}\frac{\partial L_n(\beta,1)}{\partial\lambda}\rho-\frac{1}{2}var[G(\beta)]\rho^2\right\}+o_{p\beta}(1)\\\xrightarrow{d}U&=\exp\left\{-G(\beta)\rho-\frac{1}{2}var[G(\beta)]\rho^2\right\}.\end{aligned}$$

Using the expression of the moment generating function of a normal distribution, we have $E(U)=1$ and hence (b) holds.

Asymptotic distribution. Using the results from (b), the joint distribution of

$$\left[\frac{H_{3,n}}{\sqrt{n}},\frac{H_{4,n}}{\sqrt{n}},\frac{1}{\sqrt{n}}\frac{\partial L_n(\beta,1)}{\partial\lambda},\ln\left(\frac{dP_{\beta,\lambda_n}}{dP_0}\right)\right]'$$

converges under H_0 to a Gaussian process such that

$$\mathcal{N}\left[\begin{pmatrix}0\\0\\0\\-\frac{1}{2}var[G(\beta)]\rho^2\end{pmatrix},\begin{pmatrix}V_3&0&c_3&-c_3\rho\\0&V_4&c_4&-c_4\rho\\c_3&c_4&var[G(\beta)]&-var[G(\beta)]\rho\\-c_3\rho&-c_4\rho&-var[G(\beta)]\rho&var[G(\beta)]\rho^2\end{pmatrix}\right].\tag{23}$$

Let $\omega = \kappa - \delta^2/3$ and consider

$$\begin{aligned} c_3 &= cov[h_{3i}, \partial l_i(\beta, 1)/\partial\lambda] = E[h_{3i}\partial l_i(\beta, 1)/\partial\lambda] \\ &= -\frac{1}{\sqrt{e^\omega}} E\left[\left(y_i^3 - 3y_i\right)\exp\left\{\frac{1}{2}\left[y_i^2 - \frac{(y_i+\delta)^2}{e^\omega}\right]\right\}\right], \end{aligned}$$

which follows because h_{3i} is orthogonal to both $h_{1i} = y_i$ and $h_{2i} = y_i^2 - 1$. Under H_0, $y_i \sim N(0,1)$, it follows that

$$\begin{aligned} E\left\{\left(y_i^3 - 3y_i\right)\exp\left[\frac{1}{2}\left\{y_i^2 - \frac{(y_i+\delta)^2}{e^\omega}\right\}\right]\right\} &= \frac{1}{\sqrt{2\pi}}\int\left(y^3 - 3y\right)\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{(y+\delta)^2}{2e^\omega}\right]dy \\ &= \sqrt{e^\omega}\int\left[\left(\sqrt{e^\omega}u - \delta\right)^3 - 3\left(\sqrt{e^\omega}u - \delta\right)\right]\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du \\ &= \sqrt{e^\omega}\left(-\delta^3 - 3\delta e^\omega + 3\delta\right) \end{aligned}$$

if we use the change of variable $u = (y+\delta)/\sqrt{e^\omega}$.

Hence, we have $cov[h_{3i}, \partial l_i(\beta, 1)/\partial\lambda] = \delta^3 + 3\delta(e^\omega - 1)$, and also

$$\begin{aligned} cov[h_{4i}, \partial l_i(\beta, 1)/\partial\lambda] &= E[h_{4i}\partial l_i(\beta, 1)/\partial\lambda] \\ &= -\frac{1}{\sqrt{e^\omega}} E\left[\left(y_i^4 - 6y_i^2 + 3\right)\exp\left\{\frac{1}{2}\left[y_i^2 - \frac{(y_i+\delta)^2}{e^\omega}\right]\right\}\right] \\ &= -\left[3e^{2\omega} + 6e^\omega\delta^2 + \delta^4 - 6\left(e^\omega + \delta^2\right) + 3\right] \\ &= 6\delta^2(1 - e^\omega) - \delta^4 - 3(1 - e^\omega)^2 \end{aligned}$$

by the orthogonality of the Hermite polynomials and the same change of variable as before.

Then, if we denote by $(T, \ln(U))$ the limiting joint distribution given in (23), it follows from Le Cam's third Lemma (see van der Vaart (1998)) that

$$T_n = \left[\frac{H_{3,n}}{\sqrt{n}}, \frac{H_{4,n}}{\sqrt{n}}, \frac{1}{\sqrt{n}}\frac{\partial L_n(\beta, 1)}{\partial\lambda}\right]$$

converges in distribution under $\mathcal{H}_{2n}$ to a normal distribution with mean $E(T) + cov[T, \ln(U)]$ and the same variance $V(T)$ as under H_0, which proves the following result:

Under $\mathcal{H}_{2n}$,

$$\begin{pmatrix} \frac{H_{3,n}}{\sqrt{n}} \\ \frac{H_{4,n}}{\sqrt{n}} \\ \frac{1}{\sqrt{n}}\sum_i \frac{\partial l_i}{\partial\lambda} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left[\begin{pmatrix} -c_3\rho \\ -c_4\rho \\ -var[G(\beta)]\rho \end{pmatrix}, \begin{pmatrix} V_3 & 0 & c_3 \\ 0 & V_4 & c_4 \\ c_3 & c_4 & var[G(\beta)] \end{pmatrix}\right].$$

The limiting distribution of $LM_{a,1}$ under $\mathcal{H}_{2n}$ follows from the joint distribution of $\left(\frac{H_{3,n}}{\sqrt{n}}, \frac{H_{4,n}}{\sqrt{n}}\right)$ under the same sequence of local alternatives derived above.

Finally, the limiting distribution of $LM_{b,1}^B$ test follows from the distribution of $\frac{1}{\sqrt{n}}\sum_i \frac{\partial l_i}{\partial\lambda}$ and the continuous mapping theorem.

Next, we show the results for $\mathcal{H}_{1n}$.

Contiguity. To establish the result, we need first to look at the joint distribution of $\frac{H_{3,n}}{\sqrt{n}}$, $\frac{H_{4,n}}{\sqrt{n}}$, $\frac{1}{\sqrt{n}}\frac{\partial L_n(\beta,1)}{\partial \lambda}$ and $\ln \frac{dP_{\theta_n}}{dP_0}$ under P_0. It follows from the proof of Proposition 1 that

$$\begin{pmatrix} \frac{H_{3,n}}{\sqrt{n}} \\ \frac{H_{4,n}}{\sqrt{n}} \\ \frac{1}{\sqrt{n}}\frac{\partial L_n(\beta,1)}{\partial \lambda} \\ \ln \frac{dP_{\theta_n}}{dP_0} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ -\frac{V_3 w_1^2}{2} - \frac{V_4 w_2^2}{2} \end{pmatrix}, \begin{pmatrix} V_3 & 0 & c_3 & V_3 w_1 \\ 0 & V_4 & c_4 & V_4 w_2 \\ c_3 & c_4 & var[G(\beta)] & c_3 w_1 + c_4 w_2 \\ V_3 w_1 & V_4 w_2 & c_3 w_1 + c_4 w_2 & V_3 w_1^2 + V_4 w_2^2 \end{pmatrix}\right]$$

under P_0. Contiguity is established using Le Cam's first lemma as above.

Asymptotic distribution. It follows from Le Cam's third lemma that

$$\begin{pmatrix} \frac{H_{3,n}}{\sqrt{n}} \\ \frac{H_{4,n}}{\sqrt{n}} \\ \frac{1}{\sqrt{n}}\frac{\partial L_n(\beta,1)}{\partial \lambda} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left[\begin{pmatrix} V_3 w_1 \\ V_4 w_2 \\ c_3 w_1 + c_4 w_2 \end{pmatrix}, \begin{pmatrix} V_3 & 0 & c_3 \\ 0 & V_4 & c_4 \\ c_3 & c_4 & var[G(\beta)] \end{pmatrix}\right]$$

under $\mathcal{H}_{1n}$. Therefore,

$$LM_{a,1} = \frac{H_{3,n}^2}{nV_3} + \frac{H_{4,n}^2}{nV_4} \xrightarrow{d} \chi_2^2\left(V_3 w_1^2 + V_4 w_2^2\right).$$

Again, the continuous mapping theorem establishes the asymptotic distribution of $LM_{b,1}^B$.

Finally, we turn our attention to (d). Given any $\varepsilon > 0$, consider the sets $\{|LR_n - LM_n| > \varepsilon\}$ where LR_n, LM_n correspond to either the pair $LR_{a,j}$, $LM_{a,j}$ or the pair $LR_{b,j}^B$, $LM_{b,j}^B$. Proposition 1 states the asymptotic equivalence of $LR_{a,j}$ and $LM_{a,j}$ under the null, while Proposition 2 establishes the analogous asymptotic equivalence of $LR_{b,j}^B$and $LM_{b,j}^B$. Hence, we will have that $P_0(\{|LR_n - LM_n| > \varepsilon\}) \to 0$ as $n \to \infty$. Then, the definition of contiguity implies that the same probabilities go to zero under contiguous alternatives. Thus, result (d) is a consequence of the fact that P_{θ_n} is contiguous with respect to P_0 and P_{β,λ_n} with respect to P_0, as shown in (a) and (b). □

Proof of Proposition 6

The detailed steps are analogous to the ones in Proposition 1 and can be found in Supplemental Appendix C. □

References

Amengual, D., Carrasco, M. and Sentana, E. (2020): “Testing distributional assumptions using a continuum of moments”, *Journal of Econometrics* 218, 655–689.

Amengual, D., Sentana, E. and Tian, Z. (2022): “Gaussian rank correlation and regression”, in A. Chudik, C. Hsiao and A. Timmermann (eds.) *Essays in honor of M. Hashem Pesaran: Panel Modeling, Micro Applications and Econometric Methodology*, *Advances in Econometrics* 43B, 269–306, Emerald.

Andrews, D.W.K. (1994): “Empirical process methods in econometrics”,in Engle, R.F. and McFadden, D.L. (eds.), *Handbook of Econometrics* 4, 2247–2294, Elsevier

Andrews, D.W.K. (2001): “Testing when a parameter is on the boundary of the maintained hypothesis”, *Econometrica* 69, 683–734.

Azaïs, J.-M., Gassiat, E. and Mercadier, C. (2006): “Asymptotic distribution and local power of the log-likelihood ratio test for mixtures: bounded and unbounded cases”, *Bernoulli* 12, 775–799.

Bai, J. and Ng, S. (2001): “A consistent test for conditional symmetry in time series models”, *Journal of Econometrics* 103, 225–258.

Bai, J. and Ng, S. (2005): “Tests for skewness, kurtosis, and normality for time series data”, *Journal of Business & Economic Statistics* 23, 49–60.

Berndt, E.R. (1991): *The Practice of Econometrics: Classic and Contemporary*, Addison-Wesley.

Berry, S., Carnall, M. and Spiller, P. (2006): “Airline hubbing, costs and demand”, in D. Lee (ed.), *Advances in Airline Economics Volume 1: Competition Policy and Anti-Trust*, 183–214. Elsevier.

Billingsley, P. (1968): *Convergence of Probability Measures*, John Wiley & Sons, New York.

Bonhomme, S. and Manresa, E. (2005): “Grouped patterns of heterogeneity in panel data”, *Econometrica* 83, 1147–1184.

Bontemps, C. and Meddahi, N. (2005): “Testing normality: A GMM approach”, *Journal of Econometrics* 124, 149–186.

Berry, S. and Tamer, E. (2006): "Identification in models of oligopoly entry", in R. Blundell,W. K. Newey and T. Persson (eds.), *Advances in Economics and Econometrics*. Ninth World Congress, Volume 2, 46–85. Cambridge University Press.

Carrasco, M., Hu, L. and Ploberger, W. (2014): "Optimal test for Markov switching", *Econometrica* 82, 765–784.

Chen, H. and Chen, J. (2001): "The likelihood ratio test for homogeneity in finite mixture models", *Canadian Journal of Statistics* 29, 201–215.

Chen, H., Chen, J. and Kalbfleisch, J.D. (2004): "Testing for a finite mixture model with two components", *Journal of the Royal Statistical Society* Series B 66, 95–115.

Chen, J. and Li, P. (2009): "Hypothesis Tests for Normal Mixture Models: The EM Approach", *Annals of Statistics* 37, 2523–2542.

Chen, X., Ponomareva, M. and Tamer, E. (2014): "Likelihood inference in some finite mixture models", *Journal of Econometrics* 182, 87–99.

Cho, J.S. and White, H. (2007): "Testing for regime switching", *Econometrica* 75, 1671–1720.

Compiani, G. and Kitamura, Y. (2016): "Using mixtures in econometric models: a brief review and some new results", *Econometrics Journal* 19, C95–C127.

Dovonon, P. and Renault, E. (2013): "Testing for common conditionally heteroskedastic factors", *Econometrica* 81, 2561–2586.

Dufour, J.-M., Farhat, A., Gardiol, L. and Khalaf, L. (1998): "Simulation-based finite sample normality tests in linear regressions", *Econometrics Journal* 1, 154–173

Fiorentini, G. and Sentana, E. (2007): "On the efficiency and consistency of likelihood estimation in multivariate conditionally heteroskedastic dynamic regression models", CEMFI Working paper 0713.

Fiorentini, G. and Sentana, E. (2021): "New testing approaches for mean–variance predictability", *Journal of Econometrics* 222, 516–538.

Ghosh, J.K. and Sen, P.K. (1985): "On the asymptotic performance of the loglikelihood ratio statistic for the mixture model and related results", in L. Le Cam and R. Olshen (eds.), *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* vol. 2, 789–806, Wadsworth.

Griliches, Z. and Mason, W.M. (1972): "Education, income, and ability", *Journal of Political Economy* 80, S74–S103.

Hathaway, R.J. (1985): "A constrained formulation of maximum-likelihood estimation for normal mixture distributions", *Annals of Statistics* 13, 795–800.

Heckman, J.J. and Singer, B. (1984): "A method for minimizing the impact of distributional assumptions in econometric models of duration data", *Econometrica* 52, 271–320.

Horowitz, J. and Manski, C. (1995): "Identification and Robustness with Contaminated and Corrupted Data", *Econometrica* 63, 281–302.

Jarque, C.M. and Bera, A.K. (1980): "Efficient tests for normality, homoscedasticity and serial independence of regression residuals", *Economics Letters* 6, 255–259.

Keane, M.P. and Wolpin, K.I. (1997): "The career decisions of young men", *Journal of Political Economy* 105, 473–522.

Kasahara, H. and Shimotsu, K. (2015): "Testing the number of components in normal mixture regression models", *Journal of the American Statistical Association* 110, 1632–1645.

Kiefer, N.M. and Salmon, M. (1983): "Testing normality in econometric models", *Economics Letters* 11, 123–127.

Lee, L.F. and Chesher, A. (1986): "Specification testing when score test statistics are identically zero", *Journal of Econometrics* 31, 121–149.

Mincer, J. (1974): "Schooling, experience and earnings", in *Education, Income, and Human Behavior*, 71–94, NBER.

Newey, W.K. and McFadden, D.L. (1994): "Large sample estimation and hypothesis testing", in R.F. Engle and D.L. McFadden (eds.) *Handbook of Econometrics* vol. IV, 2111–2245, Elsevier.

Neyman, J. (1937): "Smooth test for goodness of fit", *Scandinavian Actuarial Journal* 3-4, 149–199.

Pearson, K. (1894): "Contributions to the mathematical theory of evolution", *Philosophical Transactions of the Royal Society of London A* 185, 71–110.

Pollard, D. (1990): *Empirical Processes: Theory and Applications*, Institute of Mathematical Statistics.

Qu, Z. and Fan, Z. (2021): “Likelihood ratio-based tests for Markov switching”, *Review of Economic Studies* 88, 937–968.

Quandt, R.E. and Ramsey, J.B. (1978): “Estimating mixtures of normal distributions and switching regressions”, *Journal of the American Statistical Association* 73, 730–738.

van der Vaart, A.W. (1998): *Asymptotic Statistics*, Cambridge University Press.

Table 1: Bootstrap-based size corrected rejection rates at 5% significance levels for Gaussian null hypothesis.

				Score-type tests							Likelihood ratio test			Other tests			
				Θ_1		Θ_2			Θ_3								
δ	$\varkappa$	$E(x^3)$	$E(x^4)$	sup	JB	sup	LM_a	LM_b	sup	LM_a	Θ_1	Θ_2	Θ_3	CL	KaSh	KS	ACS
				$n = 500$													
		$\lambda = .975$															
2	0	-0.2	3.3	33.4	36.9	32.1	28.3	32.1	29.7	30.9	25.8	26.9	30.6	25.2	25.2	12.7	21.6
3	-3	-0.3	3.4	81.1	80.0	64.4	74.7	64.4	53.0	43.3	78.8	81.7	56.6	78.4	79.0	44.9	75.9
0	1.5	0.0	3.8	52.9	51.8	51.6	22.3	51.6	56.1	55.8	35.8	30.1	54.4	34.3	31.4	9.4	22.9
		$\lambda = .75$															
-0.4	-1.1	-0.2	3.3	37.5	38.0	25.4	28.0	25.4	30.7	30.4	34.1	48.8	31.6	41.6	40.9	32.6	32.4
0.9	0.5	-0.3	3.4	65.9	71.4	63.6	71.3	63.6	38.7	40.3	41.4	63.9	66.0	66.4	66.9	68.5	40.8
0.0	1.0	0.0	3.8	64.5	68.5	51.5	17.0	51.6	75.4	76.9	21.2	51.6	38.7	57.3	57.4	36.3	77.7
				$n = 125$													
		$\lambda = .975$															
2	0	-0.2	3.3	15.5	16.2	14.9	14.0	14.9	14.6	15.4	10.6	9.5	14.9	9.3	9.6	6.2	10.1
3	-3	-0.3	3.4	27.3	25.4	18.1	26.9	18.1	17.5	15.0	23.0	23.7	18.0	22.0	22.4	12.7	26.4
0	1.5	0.0	3.8	23.9	24.2	23.3	17.5	23.3	24.2	25.3	15.4	11.8	24.7	13.4	12.7	6.5	12.6
		$\lambda = .75$															
-0.4	-1.1	-0.2	3.3	15.9	15.7	12.4	13.5	12.4	14.2	14.4	10.5	16.5	11.8	13.5	14.0	12.0	14.8
0.9	0.5	-0.3	3.4	25.3	27.0	23.9	27.9	23.9	18.1	18.9	13.1	22.5	21.8	19.7	20.2	21.1	18.8
0.0	1.0	0.0	3.8	30.3	31.9	26.2	16.7	26.2	35.3	36.3	9.1	20.5	14.7	18.2	18.4	11.5	36.6

Notes: Results based on 10,000 replications with critical values computed using 10,000 simulated values under the null. λ denotes the mixing probability, δ the difference in means and $\varkappa$ the ratio of variances of the mixture of two normals. For both, the score-type tests and the likelihood ratio test, the three different parameter spaces are

$$\Theta_1' = [-\bar{\delta}, \bar{\delta}] \times [-\bar{\varkappa}, \bar{\varkappa}] \times [1/2, 1], \quad \Theta_2' = [-\bar{\delta}, \bar{\delta}] \times \{0\} \times [1/2, 1], \quad \text{and} \quad \Theta_3' = \{0\} \times [-\bar{\varkappa}, \bar{\varkappa}] \times [1/2, 1].$$

LM's and LR's are defined in Section 3. CL and KaSh refer to the testing procedures proposed in Chen and Li (2009) and Kasahara and Shimotsu (2015), respectively, while KS denotes the Kolmogorov-Smirnov test and ACS the CGMM test proposed in Amengual, Carrasco and Sentana (2020) with Tikhonov regularization parameter α =.01 and scale parameter $\omega^2 = 1$.

Table 2: Application to Mincer equations

Specification		Men & women		Men only		Women only	
n		534		245		289	
Skewness		-0.08		0.49		-0.56	
Kurtosis		4.72		4.68		4.70	
		Testing procedures					
		statistic	p-value	statistic	p-value	statistic	p-value
Θ_1	LM_1	751.0	.00	522.3	.00	1,234.5	.00
	JB	61.9	.00	34.2	.00	45.0	.00
	LR_1	10.7	.01	10.0	.01	11.1	.01
Θ_2	LM_2	534.1	.00	468.9	.00	963.8	.00
	$LM_{a,2}$	0.6	.62	8.8	.00	13.7	.00
	$LM_{b,2}$	534.1	.00	468.9	.00	963.8	.00
	LR_2	5.2	.07	6.2	.05	7.1	.03
Θ_3	LM_3	714.1	.00	207.9	.00	464.1	.00
	$LM_{a,3}$	61.3	.00	25.5	.00	31.3	.00
	LR_3	10.6	.00	5.0	.00	5.4	.00
JB skew		0.6	.44	8.8	.00	13.7	.00
JB kurt		61.3	.00	25.5	.00	31.3	.00

Notes: CPS85 dataset provided by the Berndt (1991). For both, the score-type tests and the likelihood ratio test, the three different parameter spaces are

$$\Theta'_1 = [-\bar{\delta}, \bar{\delta}] \times [-\bar{\varkappa}, \bar{\varkappa}] \times [1/2, 1]\,, \quad \Theta'_2 = [-\bar{\delta}, \bar{\delta}] \times \{0\} \times [1/2, 1]\,, \quad \text{and} \quad \Theta'_3 = \{0\} \times [-\bar{\varkappa}, \bar{\varkappa}] \times [1/2, 1]\,.$$

LM's and LR's are defined in Section 3. JB skew (JB kurt) refers to the Jarque-Bera skewness (kurtosis) component of the Jarque-Bera (1980) test.

Figure 1: Exclusion zones of the parameter space

Figure 1a: $H_{01} : (\delta, \varkappa) = 0$ with $\lambda \leq 1 - \varepsilon < 1$

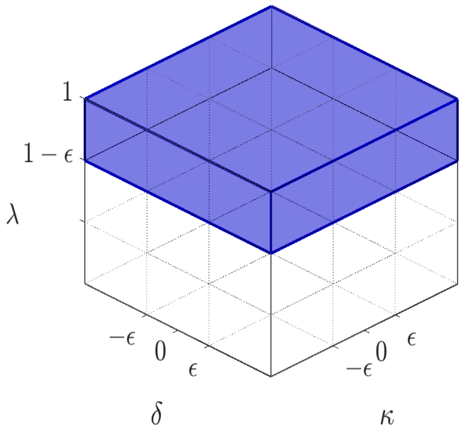


Figure 1b: $H_{02} : \lambda = 1$ with $\min\{|\delta|, |\varkappa|\} \geq \epsilon > 0$

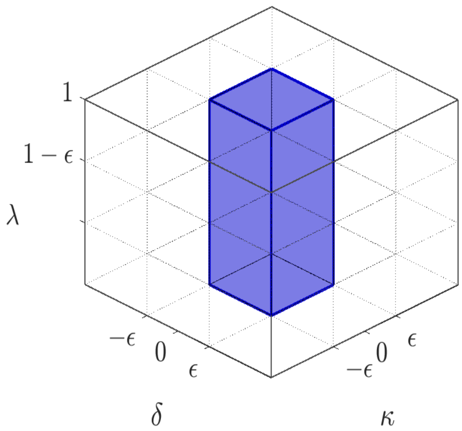


Figure 1c: Corner case, $\min\{|\delta|, |\varkappa|\} < \epsilon$, $\lambda > 1 - \epsilon$

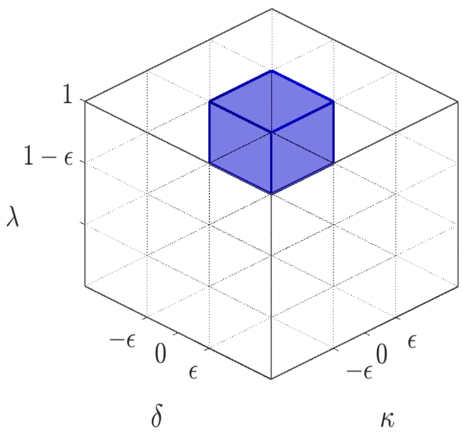


Figure 2: Reparametrization, null hypotheses and parameter space partition

Figure 2a: Transformed parameter space

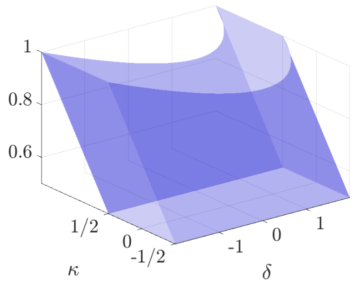


Figure 2b: The null hypothesis $H_0 : (\delta, \kappa) = 0$

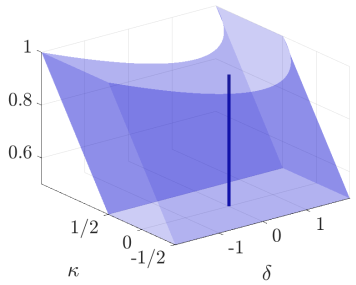


Figure 2c: The null hypothesis $H_0 : \lambda = 1$

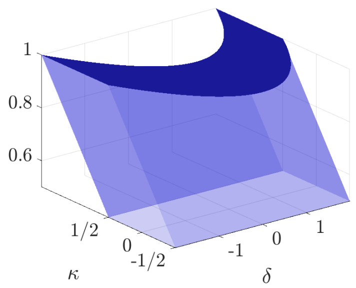


Figure 2d: The joint null hypothesis

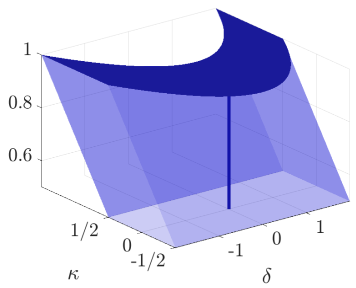


Figure 2e: Partition of the parameter space

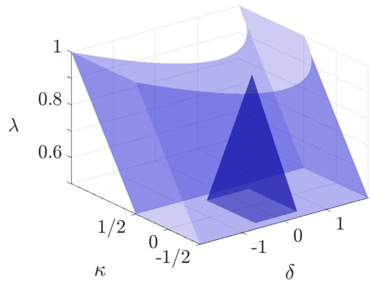


Figure 2f: The null hypothesis $H_0 : (\delta, \kappa) = 0$

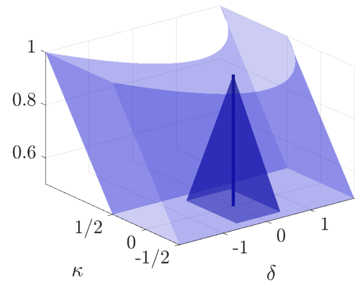


Figure 2g: The null hypothesis $H_0 : \lambda = 1$

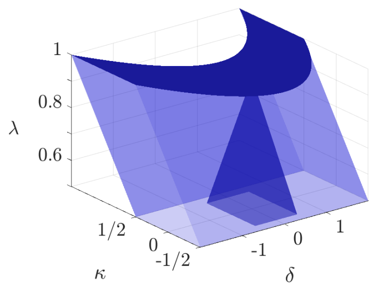


Figure 2h: The (cylindric) pyramidion

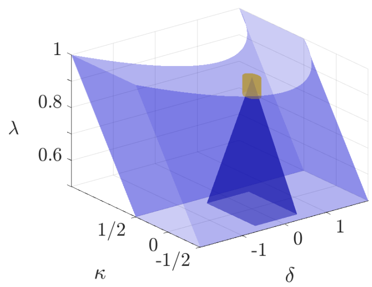


Figure 3: Unrestricted and restricted alternatives

Figure 3a: $\delta = 3$, $\varkappa = -3$ and $\lambda = 0.5$

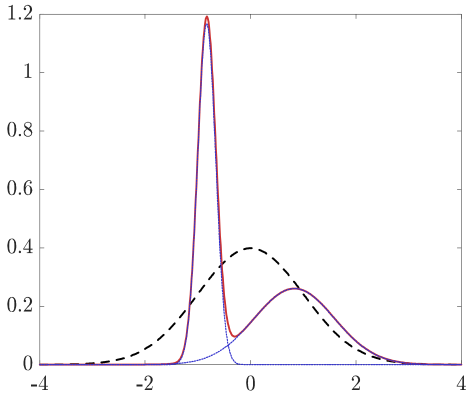


Figure 3d: $LM_{b,1}$ vs $LM_{a,1}$

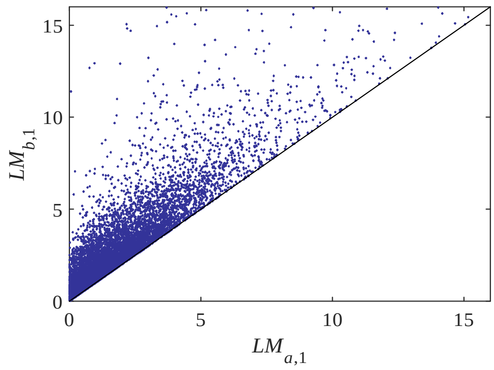


Figure 3b: $\delta = 2$, $\varkappa = 0$ and $\lambda = 0.6$

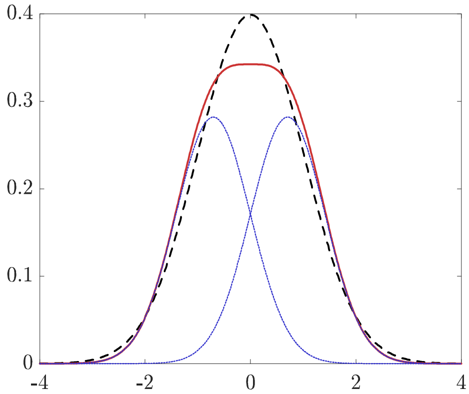


Figure 3d: $LM_{b,1}$ vs $LM_{a,1}$

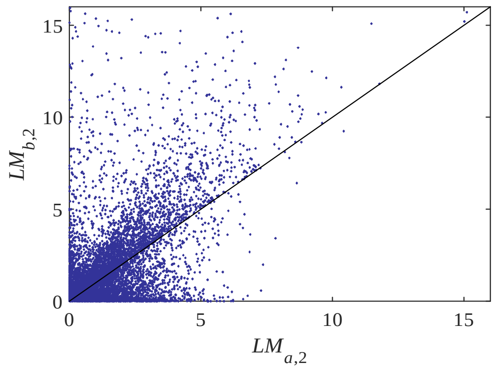


Figure 3c: $\delta = 0$, $\varkappa = 3/2$ and $\lambda = 0.5$

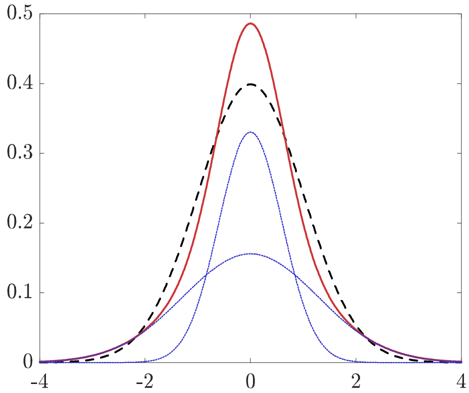
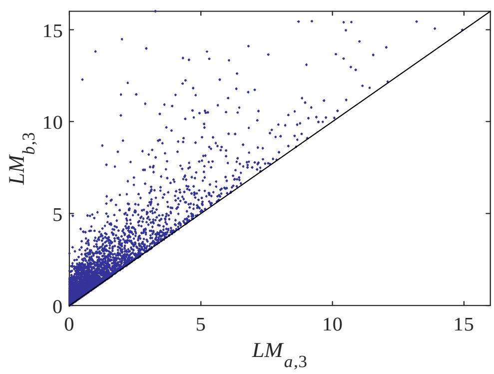


Figure 3f: $LM_{b,3}$ vs $LM_{a,3}$



Notes: (a)-(c) The dashed (black) line represents the pdf of a standard normal distribution. The continuous (red) line represents the density of the standardized Gaussian mixture described in Section 2, while the dotted (blue) lines the probability weighted densities of its components; (d)-(f) Scatter plots of the two test statistics for the three possible null hypotheses based on 10,000 replications under normality.

Figure 4: Distributions under different alternatives

Figure 4a: $\delta = 3$, $\varkappa = -3$ and $\lambda = 0.975$

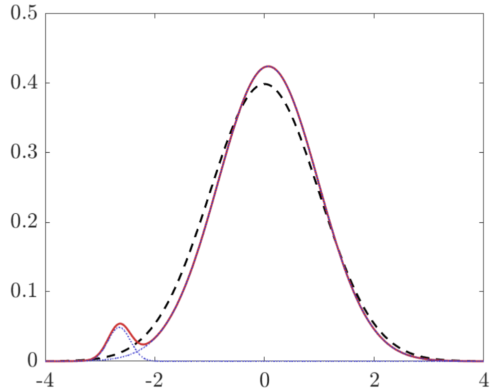


Figure 4b: $\delta = 2$, $\varkappa = 0$ and $\lambda = 0.975$

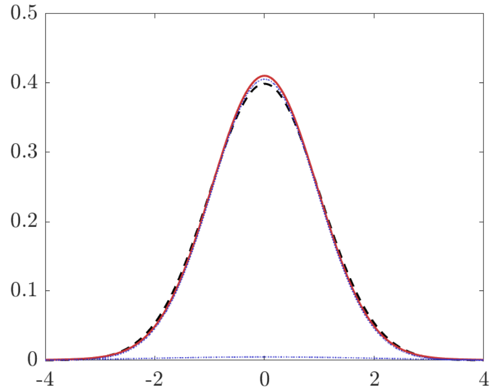


Figure 4c: $\delta = 0$, $\varkappa = 3/2$ and $\lambda = 0.975$

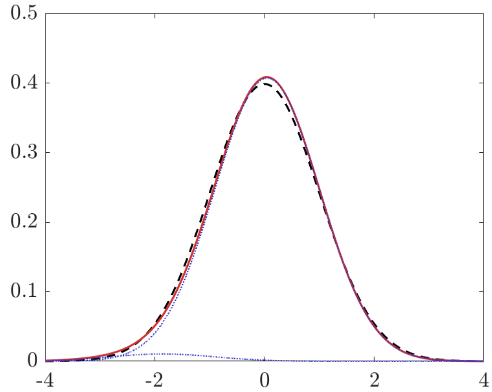


Figure 4d: $\delta = 3$, $\varkappa = -3$ and $\lambda = 0.75$

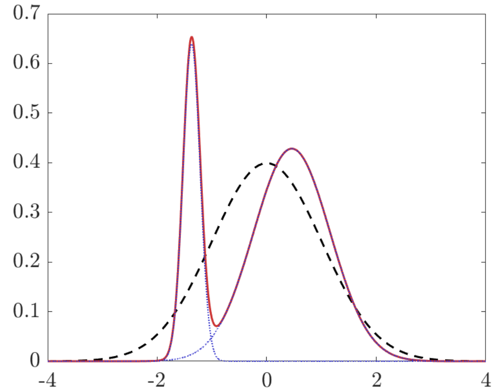


Figure 4e: $\delta = 2$, $\varkappa = 0$ and $\lambda = 0.75$

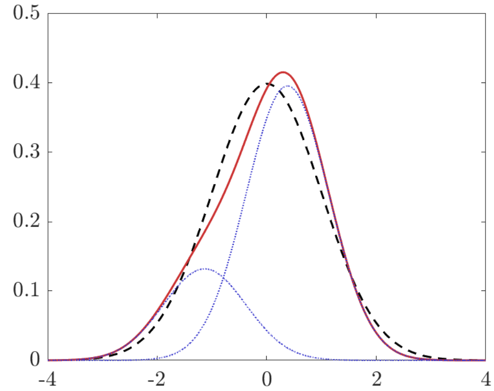
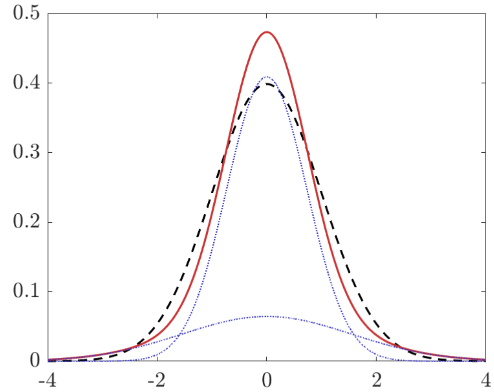


Figure 4f: $\delta = 0$, $\varkappa = 3/2$ and $\lambda = 0.75$



Notes: The dashed (black) line represents the pdf of a standard normal distribution. The continuous (red) line represents the density of the standardized Gaussian mixture described in Section 2, while the dotted (blue) lines represent the probability weighted densities of its components.

Supplemental Appendices for

Score-type tests for normal mixtures

Dante Amengual
CEMFI
<amengual@cemfi.es>

Xinyue Bei
Duke University
<xinyue.bei@duke.edu>

Marine Carrasco
Université de Montréal
<marine.carrasco@umontreal.ca>

Enrique Sentana
CEMFI
<sentana@cemfi.es>

November 2023

A Detailed proof of Proposition 1

We follow the steps outlined in the appendix of the paper.

Step 1

We want to show that for all sequences $\theta_n = (\delta_n, \kappa_n, \lambda_n) \in \Theta$ with $(\delta_n, \kappa_n) \xrightarrow{p} 0$, we have

$$LR_n(\theta_n) = LM_n^a(\theta_n) + o_p[h_n(\theta_n)], \tag{A1}$$

where $h_n(\theta) = \max\left\{1, n(1-\lambda_n)^2\delta_n^8, n(1-\lambda_n)^2\delta_n^2\kappa_n^2, n(1-\lambda_n)^2\kappa_n^4\right\}$.

Let l denote the log likelihood of the observable y, $h_3 = y(y^2-3)$ and $h_4 = y^4 - 6y^2 + 3$. The scores and relevant higher-order derivatives with respect to δ and κ at the point $(0, 0, \lambda_n)$ are

$$\frac{\partial l}{\partial \delta} = 0, \quad \frac{\partial l}{\partial \kappa} = 0,$$

$$\frac{\partial^2 l}{\partial \delta^2} = 0, \quad \frac{\partial^2 l}{\partial \delta \partial \kappa} = -\frac{1}{2}(1-\lambda_n)\lambda_n h_3, \quad \frac{\partial^2 l}{\partial \kappa^2} = \frac{1}{4}(1-\lambda_n)\lambda_n h_4,$$

$$\frac{\partial^3 l}{\partial \delta^3} = 0 \quad \text{and} \quad \frac{\partial^4 l}{\partial \delta^4} = -\frac{2}{3}(1-\lambda_n)\lambda_n(1-\lambda_n+\lambda_n^2)h_4.$$

Let

$$L_n^{[k_1,k_2]} = \frac{1}{k_1!k_2!}\frac{\partial^{k_1+k_2}L_n(\theta)}{\partial\delta^{k_1}\partial\kappa^{k_2}}\bigg|_{(0,0,\lambda_n)}$$

and

$$\triangle_n^{[k_1,k_2]} = \frac{1}{k_1!k_2!}\frac{\partial^{k_1+k_2}L_n(\theta)}{\partial\delta^{k_1}\partial\kappa^{k_2}}\bigg|_{(\tilde{\delta}_n,\tilde{\kappa}_n,\lambda_n)}$$

with $(\tilde{\delta}_n, \tilde{\kappa}_n)$ between 0 and (δ_n, κ_n). Then, taking an eighth-order Taylor expansion we get

$$\begin{aligned}
\frac{1}{2}LR_n(\theta_n) =& L_n(\theta_n) - L_n(0,0,\lambda_n) \\
=& \sqrt{n}\delta_n^4\left(A_{1n} + \delta_n A_{2n} + \sqrt{n}\delta_n^4 A_{3n}\right) \\
&+ \sqrt{n}\kappa_n^2\left[A_{4n} + \kappa_n A_{5n} + \sqrt{n}\kappa_n^2\left(A_{6n} + \kappa_n A_{7n}\right)\right] \\
&+ \sqrt{n}\delta_n\kappa_n\left[A_{8n} + \delta_n\left(A_{9n} + \sqrt{n}\delta_n^4 A_{10n}\right) + \kappa_n\left(A_{11n} + \sqrt{n}\kappa_n^2 A_{12n}\right)\right] \\
&+ n\delta_n^2\kappa_n^2\left(A_{13n} + A_{14n}\right) + \sum_{j+k=9}\frac{1}{n}\Delta^{[j,k]}n\delta_n^j\kappa_n^k,
\end{aligned} \tag{A2}$$

where

$$A_{1n} = \left\{\frac{1}{\sqrt{n}}L_n^{[4,0]}\right\}, \quad A_{2n} = \sum_{j=5}^{7}\left\{\frac{1}{\sqrt{n}}L_n^{[j,0]}\right\}\delta_n^{i-5}, \quad A_{3n} = \left\{\frac{1}{n}L_n^{[8,0]}\right\}, \quad A_{4n} = \left\{\frac{1}{\sqrt{n}}L_n^{[0,2]}\right\},$$

$$A_{5n} = \left\{\frac{1}{\sqrt{n}}L_n^{[0,3]}\right\}, \quad A_{6n} = \frac{1}{n}L_n^{[0,4]}, \quad A_{7n} = \sum_{j=5}^{8}\left\{\frac{1}{n}L_n^{[0,j]}\right\}\kappa_n^{j-5}, \quad A_{8n} = \left\{\frac{1}{\sqrt{n}}L_n^{[1,1]}\right\},$$

$$A_{9n} = \sum_{j=2}^{5} \left\{ \frac{1}{\sqrt{n}} L_n^{[j,1]} \right\} \delta_n^{j-2}, \quad A_{10n} = \sum_{j=6}^{7} \left\{ \frac{1}{n} L_n^{[j,1]} \right\} \delta_n^{j-6}, \quad A_{11n} = \sum_{j=2}^{3} \left\{ \frac{1}{\sqrt{n}} L_n^{[1,j]} \right\} \kappa_n^{j-2},$$

$$A_{12n} = \sum_{j=4}^{7} \left\{ \frac{1}{n} L_n^{[1,j]} \right\} \kappa_n^{j-4}, \quad A_{13n} = \frac{1}{n} L_n^{[2,2]} \text{ and } A_{14n} = \sum_{\substack{8\geq j+k\geq 5 \\ j\geq 2, k\geq 2}} \left\{ \frac{1}{n} L_n^{[j,k]} \right\} \delta_n^{j-2} \kappa_n^{k-2}.$$

Next, we have to show that

$$\sum_{j+k=9} \Delta^{[j,k]} \delta_n^j \kappa_n^k = o_p[h_n(\theta_n)]. \tag{A3}$$

To do so, it is worth noticing that for $j + k = 9$,

$$\left| \frac{1}{n} \Delta_n^{[j,k]} \right| \leq \left| \frac{1}{n} \frac{1}{j!k!} \left. \frac{\partial^{j+k} L_n(\theta)}{\partial \delta^j \partial \kappa^k} \right|_{(0,0,\lambda_n)} \right| + \left| \frac{1}{n} \frac{1}{j!k!} \left. \frac{\partial^{j+k+1} L_n(\theta)}{\partial \delta^{j+1} \partial \kappa^k} \right|_{(\bar{\delta}_n, \bar{\kappa}_n, \lambda_n)} \right| \left| \tilde{\delta}_n \right| \tag{A4}$$

$$+ \left| \frac{1}{n} \frac{1}{j!k!} \left. \frac{\partial^{j+k+1} L_n(\theta)}{\partial \delta^j \partial \kappa^{k+1}} \right|_{(\bar{\delta}_n, \bar{\kappa}_n, \lambda_n)} \right| |\tilde{\kappa}_n|$$

$$\leq \left| \frac{1}{j!k!} \left[E \left. \frac{\partial^{j+k} l(\theta)}{\partial \delta^j \partial \kappa^k} \right|_{(0,0,\lambda_n)} \right] + O_p\left(\frac{1}{\sqrt{n}} \right) \right| \tag{A5}$$

$$+ (1 - \lambda_n) \frac{1}{j!k!}$$

$$\times \left\{ \left| E \left[\left. \frac{\partial^{j+k+1} l(\theta)}{\partial \delta^{j+1} \partial \kappa^k} \right|_{(0,0,\lambda_n)} \right] \right| + \left| \left[E \left. \frac{\partial^{j+k+1} L_n(\theta)}{\partial \delta^j \partial \kappa^{k+1}} \right|_{(0,0,\lambda_n)} \right] \right| + o_p(1) \right\}$$

$$= O\left[(1 - \lambda_n)^2 \right] + O_p\left(\frac{1}{\sqrt{n}} \right) + o_p(1 - \lambda_n), \tag{A6}$$

where (A4) comes from the mean-value theorem, (A5) follows from the central limit theorem and

$$\max\{ \left| \tilde{\delta}_n \right|, |\tilde{\kappa}_n| \} \leq \max\{ |\delta_n|, |\kappa_n| \} \leq (1 - \lambda_n),$$

while (A6) follows from

$$E \left[\left. \frac{\partial^{j'+k'} l(\theta)}{\partial \delta^{j'} \partial \kappa^{k'}} \right|_{(0,0,\lambda_n)} \right] = O[(1 - \lambda_n)^2],$$

for $j' + k' = 9$ and $j' + k' = 10$, which can be easily checked by hand. Then,

$$\sum_{j+k=9} \Delta^{[j,k]} \delta_n^j \kappa_n^k = \sum_{j+k=9} \left\{ O\left[(1 - \lambda_n)^2 \right] + O_p\left(\frac{1}{\sqrt{n}} \right) + o_p\left[(1 - \lambda_n) \right] \right\} n \delta_n^j \kappa_n^k$$

$$= \sum_{j+k=9} O\left[(1 - \lambda_n)^2 \right] n \delta_n^j \kappa_n^k + \sum_{j+k=9} O_p(\sqrt{n} \delta_n^j \kappa_n^k) + \sum_{j+k=9} o_p\left[(1 - \lambda_n) \right] n \delta_n^j \kappa_n^k$$

$$= o_p\left[h_n(\theta_n) \right],$$

which follows from δ_n, $\kappa_n = o_p(1)$ and $(1 - \lambda_n) \geq \max\{|\delta_n|, |\kappa_n|\}$.

If we then use (A2) and (A3), we can show that

$$\frac{1}{2}LR_n(\theta_n) = \sqrt{n}\delta_n^4\left(A_{1n}+\sqrt{n}\delta_n^4 A_{3n}\right)+\sqrt{n}\kappa_n^2\left(A_{4n}+\sqrt{n}\kappa_n^2 A_{6n}\right) + \sqrt{n}\delta_n\kappa_n\left(A_{8n}+\sqrt{n}\delta_n\kappa_n A_{13n}\right)+o_p[h_n(\theta_n)], \tag{A7}$$

which follows from the fact that A_{1n} to A_{13n} are $O_p(1)$, and $A_{14n} = o_p(1)$ because the terms in curly brackets are $O_p(1)$. Also,

$$\begin{aligned}\frac{1}{2}LR_n(\theta_n) =& -\frac{\lambda_n(1-\lambda_n+\lambda_n^2)}{36}\frac{H_{4,n}}{\sqrt{n}}\sqrt{n}(1-\lambda_n)\delta_n^4 \\ &-\frac{1}{2}\left[\frac{\lambda_n(1-\lambda_n+\lambda_n^2)}{36}\right]^2 V_4 n(1-\lambda_n)^2\delta_n^8 \\ &+\frac{\lambda_n}{8}\frac{H_{4,n}}{\sqrt{n}}\sqrt{n}(1-\lambda_n)\kappa_n^2-\frac{1}{2}\left(\frac{\lambda_n}{8}\right)^2 V_4 n(1-\lambda_n)^2\kappa_n^4 \\ &-\frac{\lambda_n}{2}\frac{H_{3,n}}{\sqrt{n}}\sqrt{n}(1-\lambda_n)\delta_n\kappa_n-\frac{1}{2}\left(\frac{\lambda_n}{2}\right)^2 V_3 n(1-\lambda_n)^2\delta_n^2\kappa_n^2+o_p[h_n(\theta_n)] \qquad \text{(A8)} \\ =& \frac{H_{3,n}}{\sqrt{n}}w_{1n}-\frac{1}{2}V_3 w_{1n}^2+\frac{H_{4,n}}{\sqrt{n}}w_{2n}-\frac{1}{2}V_4 w_{2n}^2+o_p[h_n(\theta_n)], \qquad \text{(A9)}\end{aligned}$$

with

$$w_{1n} = -\frac{\lambda_n}{2}\sqrt{n}(1-\lambda_n)\delta_n\kappa_n \text{ and } w_{2n} = -\frac{\lambda_n(1-\lambda_n+\lambda_n^2)}{36}\sqrt{n}(1-\lambda_n)\delta_n^4+\frac{\lambda_n}{8}\sqrt{n}(1-\lambda_n)\kappa_n^2, \tag{A10}$$

where in the first step we re-write (A7) as (A8). Then, letting

$$l^{[k_1,k_2]} = \frac{1}{k_1!k_2!}\frac{\partial^{k_1+k_2}l}{\partial\delta^{k_1}\partial\kappa^{k_2}},$$

the result follows from

$$\frac{1}{n}L_n^{[8,0]} = -\frac{1}{2}E[(l^{[4,0]})^2]+O_p(n^{-\frac{1}{2}}) \text{ and } \frac{1}{n}L_n^{[0,4]} = -\frac{1}{2}E[(l^{[0,2]})^2]+O_p(n^{-\frac{1}{2}}),$$

(see Lemma 1 in Rotnitzky et al (2000)), and

$$\frac{1}{n}L_n^{[2,2]} = -\frac{1}{2}E[(l^{[1,1]})^2]+O_p(n^{-\frac{1}{2}}),$$

which can easily be checked by hand. As for the second step, it is a simple rearrangement of terms to go from (A8) to (A9). Therefore, the only difference in the leading terms is the coefficient of V_4, namely,

$$w_{2n}^2-\left(\frac{\lambda_n}{8}\right)^2 n(1-\lambda_n)^2\kappa_n^4-\left[\frac{\lambda_n(1-\lambda_n+\lambda_n^2)}{36}\right]^2 n(1-\lambda_n)^2\delta_n^8 = O_p[n(1-\lambda_n)^2\delta_n^4\kappa_n^2] = o_p[h_n(\theta_n)],$$

as required.

Step 2

First, we show that $h_n(\theta_n^{LM}) = O_p(1)$. By definition, we have

$$LM_n^a(\theta) = 2\frac{1}{\sqrt{n}}H_{3,n}w_1 + 2\frac{1}{\sqrt{n}}H_{4,n}w_2 - V_3w_1^2 - V_4w_2^2$$
$$= -V_3\left(w_1 - \frac{1}{V_3}\frac{H_{3,n}}{\sqrt{n}}\right)^2 + \frac{1}{V_3}\left(\frac{H_{3,n}}{\sqrt{n}}\right)^2 - V_4\left(w_2 - \frac{1}{V_4}\frac{H_{4,n}}{\sqrt{n}}\right)^2 + \frac{1}{V_4}\left(\frac{H_{4,n}}{\sqrt{n}}\right)^2.$$

Let w_{1n}^{LM} and w_{2n}^{LM} be defined as in (A10) with $\delta_n = \delta_n^{LM}$, $\kappa_n = \kappa_n^{LM}$ and $\lambda_n = \lambda_n^{LM}$. It is straightforward to see that $w_{1n}^{LM} = O_p(1)$ and $w_{2n}^{LM} = O_p(1)$ because

$$\frac{n^{-\frac{1}{2}}H_{3,n}}{V_3} = O_p(1) \ \text{ and } \ \frac{n^{-\frac{1}{2}}H_{4,n}}{V_4} = O_p(1)$$

by the central limit theorem. Next, we have that

$$\left|\sqrt{n}(1-\lambda_n^{LM})\delta_n^{LM}\kappa_n^{LM}\right| = \left|\frac{2w_{1n}^{LM}}{\lambda_n^{LM}}\right| \leq \left|4w_{1n}^{LM}\right| = O_p(1),$$

whence

$$\sqrt{n}(1-\lambda_n^{LM})\delta_n^{LM}\kappa_n^{LM} = O_p(1). \tag{A11}$$

In addition, we also have

$$\left|\sqrt{n}(1-\lambda_n^{LM})\left(\kappa_n^{LM}\right)^2 - \frac{2[1-\lambda_n^{LM}+\left(\lambda_n^{LM}\right)^2]}{9}\sqrt{n}(1-\lambda_n^{LM})\left(\delta_n^{LM}\right)^4\right| = \left|\frac{8}{\lambda_n^{LM}}w_{2n}^{LM}\right|$$
$$\leq 16\left|w_{2n}^{LM}\right| = O_p(1).$$

Then by Lemma 5, $\sqrt{n}(1-\lambda_n^{LM})\left(\kappa_n^{LM}\right)^2 = O_p(1)$ and $\sqrt{n}(1-\lambda_n^{LM})\left(\delta_n^{LM}\right)^4 = O_p(1)$. Together with (A11), we have $h_n(\theta_n^{LM}) = O_p(1)$. Moreover, it holds that $\delta_n^{LM}, \kappa_n^{LM} = o_p(1)$ because

$$\sqrt{n}(|\kappa_n^{LM}|)^3 \leq \sqrt{n}(\kappa_n^{LM})^2(1-\lambda_n^{LM}) = O_p(1)$$

and

$$\sqrt{n}(|\delta_n^{LM}|)^5 \leq \sqrt{n}(\delta_n^{LM})^4(1-\lambda_n^{LM}) = O_p(1),$$

as desired.

Step 3

Next, we show Step 3.1: $(\delta_n^{LR}, \kappa_n^{LR}) \xrightarrow{p} 0$, and Step 3.2: $h_n(\delta_n^{LR}, \kappa_n^{LR}, \lambda_n^{LR}) = O_p(1)$.

Step 3.1

Let $l_0(\theta) = E_{(0,0,\lambda)}\left[l(\theta)\right]$. Invoking Lemma 6, we have

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}L_n(\theta) - l_0(\theta)\right| \xrightarrow{p} 0 \tag{A12}$$

(i.e. uniform convergence). Moreover, for all $\epsilon > 0$, we have that

$$l_0\left(0,0,\lambda\right) > \sup_{\delta^2+\kappa^2>\epsilon,\theta\in\mathcal{P}_a} l_0(\theta) \tag{A13}$$

(i.e. well separated maximum), which follows from the fact that $\delta = \kappa = 0$ is the unique maximizer (note that $(1-\lambda) \geq \max\{|\delta|,|\kappa|\}$), $l_0(\theta)$ is continuous, and Θ is compact. Hence, we have that $(\delta_n^{LR}, \kappa_n^{LR}) = o_p(1)$ by virtue of Lemma A1 in Andrews (1993).

Step 3.2

$h_n(\theta_n^{LR}) = O_p(1)$ follows directly from Step 3.2.1 and Step 3.2.2 below.

Step 3.2.1

We first show that $n\left(1-\lambda_n^{LR}\right)^2\left(\delta_n^{LR}\right)^8 = O_p(1)$ and $n\left(1-\lambda_n^{LR}\right)^2\left(\kappa_n^{LR}\right)^4 = O_p(1)$. By contradiction, assume that either $n\left(1-\lambda_n^{LR}\right)^2\left(\delta_n^{LR}\right)^8 \neq O_p(1)$ or $n\left(1-\lambda_n^{LR}\right)^2\left(\kappa_n^{LR}\right)^4 \neq O_p(1)$, so that there exists $\epsilon > 0$ such that for all M it holds that $\Pr(A_n) > \epsilon$ i.o., where

$$A_n = \left\{\frac{1}{288}n^{\frac{1}{2}}\left(1-\lambda_n^{LR}\right)\left(\delta_n^{LR}\right)^4 > M\right\} \cup \left\{\frac{1}{144}n^{\frac{1}{2}}\left(1-\lambda_n^{LR}\right)\left(\kappa_n^{LR}\right)^2 > M\right\}.$$

Since $H_{3,n}/\sqrt{n}$ and $H_{4,n}/\sqrt{n}$ are $O_p(1)$, there exists M_1 such that $\Pr(B_n) \geq 1-\epsilon/4$ for all n, where

$$B_n = \left\{\left|\frac{H_{3,n}}{\sqrt{n}}\right| < M_1\right\} \cap \left\{\left|\frac{H_{4,n}}{\sqrt{n}}\right| < M_1\right\}.$$

Next, let $r_n(\theta) = LR_n(\theta) - LM_n(\theta)$. Since κ_n^{LR}, δ_n^{LR} and $r_n(\theta_n^{LR})/h(\theta_n^{LR})$ are $o_p(1)$, with positive $\xi < 1/3$, we have that $\Pr\left(C_n\right) \geq 1-\epsilon/4$ ult., where

$$C_n = \left\{|\kappa_n^{LR}| < \xi, |\delta_n^{LR}| < \xi\right\} \cap \left\{\left|\frac{r_n(\theta_n^{LR})}{h_n(\theta_n^{LR})}\right| < \xi\left(\frac{1}{288}\right)^2\right\}.$$

Let us define w_{2n}^{LR} in the same way as w_{2n}, but with the parameters λ_n, κ_n and δ_n replaced by λ_n^{LR}, κ_n^{LR} and δ_n^{LR}, respectively. In addition, let

$$D_n = \left\{|w_{2n}^{LR}| \leq \frac{1}{288}\max\left[n^{\frac{1}{2}}\left(1-\lambda_n^{LR}\right)\left(\delta_n^{LR}\right)^4, 2n^{\frac{1}{2}}\left(1-\lambda_n^{LR}\right)\left(\kappa_n^{LR}\right)^2\right]\right\},$$

$$E_n = \{n^{\frac{1}{2}}\left(\delta_n^{LR}\right)^4 > 2n^{\frac{1}{2}}\left(\kappa_n^{LR}\right)^2\} \text{ and } F_n = \left\{|w_{2n}^{LR}| < |w_{1n}^{LR}|\right\}.$$

Then, we can show that for all M,

$$\Pr(A_n \cap B_n \cap C_n) \geq \Pr(A_n) + \Pr(B_n) + \Pr(C_n) - 2 \geq \frac{\epsilon}{2} \text{ i.o.},$$

where the first inequality follows from $\Pr(A\cap B) \geq \Pr(A)+\Pr(B)-1$, and the second inequality follows from the lower bounds of $\Pr(A_n)$, $\Pr(B_n)$ and $\Pr(C_n)$ derived above.

In addition, let $M > M_1/\xi$ and consider $A_n \cap B_n \cap C_n \cap D_n \cap E_n$. We next use Lemma 7 to show that $A_n \cap B_n \cap C_n \cap D_n \cap E_n \subset \left\{LR(\delta_n^{LR}, \kappa_n^{LR}, \lambda_n^{LR}) < 0\right\} = \emptyset$. To do so, let us check all the required conditions. First, notice that $|H_{3,n}/\sqrt{n}| < M_1$ and $|H_{4,n}/\sqrt{n}| < M_1$ are satisfied on B_n. Second, we can easily check that

$$|w_{1n}^{LR}| > \frac{M_1}{\xi} \quad \text{and } |w_{1n}^{LR}| > |w_{2n}^{LR}|$$

because

$$\begin{aligned} n\left(1-\lambda_n^{LR}\right)^2\left(\kappa_n^{LR}\delta_n^{LR}\right)^2 &= n^{\frac{1}{2}}\left(1-\lambda_n^{LR}\right)\left(\kappa_n^{LR}\right)^2 n^{\frac{1}{2}}\left(1-\lambda_n^{LR}\right)\left(\delta_n^{LR}\right)^2 \\ &= \left\{\frac{8w_{2n}^{LR}}{\lambda_n^{LR}} + \frac{2}{9}[1-\lambda_n^{LR}+\left(\lambda_n^{LR}\right)^2]n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\delta_n^{LR}\right)^4\right\} \qquad \text{(A14)} \\ &\times n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\delta_n^{LR}\right)^2 \\ &\geq \left[-16\left|w_{2n}^{LR}\right| + \frac{1}{6}n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\delta_n^{LR}\right)^4\right] n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\delta_n^{LR}\right)^2 \qquad \text{(A15)} \\ &\geq \left(\frac{1}{6}-\frac{1}{18}\right) n\left(1-\lambda_n^{LR}\right)^2\left(\delta_n^{LR}\right)^6 \geq \frac{n\left(1-\lambda_n^{LR}\right)^2\left(\delta_n^{LR}\right)^8}{9\xi^2}, \qquad \text{(A16)} \end{aligned}$$

where (A14) follows from the definition of w_{2n}^{LR}, (A15) follows from the bound of λ_n^{LR}, the first inequality of (A16) is a direct consequence of combining $D_n$with E_n, while the second one follows from the definition of C_n.

Then, we have

$$\left|w_{1n}^{LR}\right| = \frac{\lambda_n^{LR}}{2}\left|n^{\frac{1}{2}}(1-\lambda_n^{LR})\kappa_n^{LR}\delta_n^{LR}\right| \geq \frac{1}{4}\frac{n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\delta_n^{LR}\right)^4}{3\xi} \qquad \text{(A17)}$$

$$\begin{cases} \geq \frac{24M}{\xi} > \frac{M_1}{\xi} & \text{(i)} \\ > \frac{1}{288}n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\delta_n^{LR}\right)^4 \geq |w_{2n}^{LR}| & \text{(ii)} \end{cases} \qquad \text{(A18)}$$

where (A17) follows from (A16), (A18i) follows from combining A_n with E_n and $M_1 < M$, while (A18ii) follows from combining D_n with E_n.

Next, we check that $r_n(\theta_n^{LR})/\left(w_{1n}^{LR}\right)^2 < \xi$ thanks to

$$\left|n^{\frac{1}{2}}(1-\lambda_n^{LR})\kappa_n^{LR}\delta_n^{LR}\right| \geq \frac{n^{\frac{1}{2}}(1-\lambda_n^{LR})(\delta_n^{LR})^4}{3\xi} \geq n^{\frac{1}{2}}(1-\lambda_n^{LR})(\delta^{LR})^4 \qquad \text{(A19)}$$

$$\left|n^{\frac{1}{2}}(1-\lambda_n^{LR})\kappa_n^{LR}\delta_n^{LR}\right| \geq \frac{n^{\frac{1}{2}}(1-\lambda_n^{LR})(\delta_n^{LR})^4}{3\xi} \geq \frac{2n^{\frac{1}{2}}(1-\lambda_n^{LR})(\kappa^{LR})^2}{3\xi} > n^{\frac{1}{2}}(1-\lambda_n^{LR})(\kappa^{LR})^2, \qquad \text{(A20)}$$

where (A19) follows from (A16) and $\xi < 1/3$, and (A20) follows from the definition of E_n and $\xi < 1/3$. Thus, $h_n(\theta_n^{LR}) = n(1-\lambda_n^{LR})^2(\kappa_n^{LR}\delta_n^{LR})^2$ and, as a result,

$$\begin{aligned} \left|\frac{r_n(\theta_n^{LR})}{\left(w_{1n}^{LR}\right)^2}\right| &= \left|\frac{r_n(\theta_n^{LR})}{h_n(\theta_n^{LR})}\frac{h_n(\theta_n^{LR})}{\left(w_{1n}^{LR}\right)^2}\right| = \left|\frac{r_n(\theta_n^{LR})}{h_n(\theta_n^{LR})}\right|\left|\frac{n\left(1-\lambda_n^{LR}\right)^2\left(\kappa_n^{LR}\delta_n^{LR}\right)^2}{\left(w_{1n}^{LR}\right)^2}\right| \\ &< \xi\left(\frac{1}{288}\right)^2\frac{4}{[\lambda_n^{LR}]^2} < \xi, \qquad \text{(A21)} \end{aligned}$$

where (A21) follows from the definitions of C_n and w_{1n}^{LR}. But then, we have that $LR(\theta_n^{LR}) < 0$ conditional on $A_n \cap B_n \cap C_n \cap D_n \cap E_n$ by virtue of Lemma 7, and consequently, that $A_n \cap B_n \cap$

$C_n \cap D_n \cap E_n = \emptyset$.

Consider now $A_n \cap B_n \cap C_n \cap D_n \cap E_n^c$. We can use Lemma 7 again to show that $A_n \cap B_n \cap C_n \cap D_n \cap E_n^c \subset \left\{LR(\theta_n^{LR}) < 0\right\} = \emptyset$. First, notice that $|H_{3,n}/\sqrt{n}| < M_1$ and $|H_{4,n}/\sqrt{n}| < M_1$ are satisfied on B_n. Next, we have to check that $|w_{1n}^{LR}| > M_1/\xi$ and $|w_{1n}^{LR}| > |w_{2n}^{LR}|$. To do so, notice that

$$n\left(1-\lambda_n^{LR}\right)^2\left(\kappa_n^{LR}\delta_n^{LR}\right)^2 \geq n^{\frac{1}{2}}\left(\kappa_n^{LR}\right)^2 n^{\frac{1}{2}}\left(\delta_n^{LR}\right)^4 \frac{1}{\xi^2}\left(1-\lambda_n^{LR}\right)^2 \tag{A22}$$

$$\geq n^{\frac{1}{2}}\left(1-\lambda_n^{LR}\right)\left(\kappa_n^{LR}\right)^2 \frac{36}{(1-\lambda_n+\lambda_n^2)} \tag{A23}$$

$$\times\left[\frac{1}{8}\sqrt{n}\left(1-\lambda_n^{LR}\right)\left(\kappa_n^{LR}\right)^2 - \frac{w_{2n}^{LR}}{\lambda_n}\right]\frac{1}{\xi^2}$$

$$\geq n^{\frac{1}{2}}\left(1-\lambda_n^{LR}\right)\left(\kappa_n^{LR}\right)^2 36 \tag{A24}$$

$$\times\left[\frac{1}{8}\sqrt{n}\left(1-\lambda_n^{LR}\right)\left(\kappa_n^{LR}\right)^2 - 2|w_{2n}^{LR}|\right]\frac{1}{\xi^2}$$

$$\geq 4n\left(1-\lambda_n^{LR}\right)^2\left(\kappa_n^{LR}\right)^4\frac{1}{\xi^2}, \tag{A25}$$

where (A22) follows from the definition of C_n, (A23) follows from the definition of w_{2n}^{LR}, (A24) follows from the bound of λ_n^{LR}, and (A25) follows from combining D_n with E_n^c.

Then,

$$|w_{1n}^{LR}| = \left|\frac{(1-\lambda_n^{LR})\lambda_n^{LR}}{2} n^{\frac{1}{2}}\kappa_n^{LR}\delta_n^{LR}\right| \geq \frac{1}{4}\frac{2n^{\frac{1}{2}}(1-\lambda_n^{LR})(\kappa_n^{LR})^2}{\xi} > \frac{1}{72}n^{\frac{1}{2}}(1-\lambda_n^{LR})(\kappa_n^{LR})^2 \tag{A26}$$

$$\begin{cases} > M > \frac{M_1}{\xi} & \text{(i)}, \\ \geq |w_2^{LR}| & \text{(ii)}, \end{cases} \tag{A27}$$

where (A26) follows from (A25), (A27i) follows from combining A_n with E_n^c, and (A27ii) follows from combining D_n with E_n^c.

To check that $r_n(\theta_n^{LR})/\left(w_{1n}^{LR}\right)^2 < \xi$, let us write

$$\left|n^{\frac{1}{2}}(1-\lambda_n^{LR})\kappa_n^{LR}\delta_n^{LR}\right| \geq \frac{2n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\kappa_n^{LR}\right)^2}{\xi} > n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\kappa_n^{LR}\right)^2 \tag{A28}$$

$$\left|n^{\frac{1}{2}}(1-\lambda_n^{LR})\kappa_n^{LR}\delta_n^{LR}\right| \geq \frac{2n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\kappa_n^{LR}\right)^2}{\xi} > \frac{n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\delta_n^{LR}\right)^4}{\xi}$$

$$> n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\delta_n^{LR}\right)^4, \tag{A29}$$

where (A28) follows from (A25), and (A29) follows from the definition of E_n^c. Thus, $h_n(\theta_n^{LR}) = n(1-\lambda_n^{LR})^2(\kappa_n^{LR}\delta_n^{LR})^2$ and, consequently,

$$\left|\frac{r_n(\theta_n^{LR})}{\left(w_{1n}^{LR}\right)^2}\right| = \left|\frac{r_n(\theta_n^{LR})}{h_n(\theta_n^{LR})}\right|\left|\frac{n\left(1-\lambda_n^{LR}\right)^2\left(\kappa_n^{LR}\delta_n^{LR}\right)^2}{\left(w_{1n}^{LR}\right)^2}\right| = \left|\frac{r_n(\theta_n^{LR})}{h_n(\theta_n^{LR})}\right|\left|\frac{4}{(\lambda_n^{LR})^2}\right| < \xi, \tag{A30}$$

where the last inequality in (A30) follows from the definition of C_n. By Lemma 7, we have $LR(\theta_n^{LR}) < 0$ conditional on $A_n \cap B_n \cap C_n \cap D_n \cap E_n^c$, and thus, $A_n \cap B_n \cap C_n \cap D_n \cap E_n^c = \emptyset$.

Consider now the case $A_n \cap B_n \cap C_n \cap D_n^c \cap F_n$. We can use Lemma 7 once again to show that $A_n \cap B_n \cap C_n \cap D_n^c \cap F_n \subset \{LR(\theta_n^{LR}) < 0\} = \emptyset$. Noticing that $|w_{1n}^{LR}| > M > M_1/\xi$ is satisfied by combining A_n with D_n^c and F_n, and that $|w_{1n}^{LR}| > |w_{2n}^{LR}|$ is satisfied by F_n, we have to check that $|r_n(\theta_n^{LR})/\left(w_{1n}^{LR}\right)^2| < \xi$. To do so,

$$\begin{aligned} \left|\frac{r_n(\theta_n^{LR})}{\left(w_{1n}^{LR}\right)^2}\right| &= \left|\frac{r_n(\theta_n^{LR})}{h_n(\theta_n^{LR})}\right| \qquad \text{(A31)} \\ &\times \left|\frac{\max\left\{1, n\left(1-\lambda_n^{LR}\right)^2\left(\kappa_n^{LR}\right)^4, n\left(1-\lambda_n^{LR}\right)^2\left(\delta_n^{LR}\right)^8, n\left(1-\lambda_n^{LR}\right)^2\left(\kappa_n^{LR}\delta_n^{LR}\right)^2\right\}}{\left(w_{1n}^{LR}\right)^2}\right| \\ &< \left|\frac{r_n(\theta_n^{LR})}{h_n(\theta_n^{LR})}\right| \left|\frac{\max\left\{\left(288 w_{2n}^{LR}\right)^2, \left(2w_{1n}^{LR}/\lambda_n^{LR}\right)^2\right\}}{\left(w_{1n}^{LR}\right)^2}\right| \qquad \text{(A32)} \\ &\leq \left|\frac{r_n(\theta_n^{LR})}{h_n(\theta_n^{LR})}\right| (288)^2 \leq \xi, \end{aligned}$$

where (A31) to (A32) follow from the definitions of D_n^c and w_1. By Lemma 7, we have that

$$LR(\delta_n^{LR}, \kappa_n^{LR}, \lambda_n^{LR}) < 0,$$

conditional on $A_n \cap B_n \cap C_n \cap D_n^c \cap F_n$, and therefore $A_n \cap B_n \cap C_n \cap D_n^c \cap F_n = \emptyset$.

Finally, consider the case $A_n \cap B_n \cap C_n \cap D_n^c \cap F_n^c$, in which

$$\begin{aligned} \frac{h_n(\theta_n^{LR})}{\left(w_{2n}^{LR}\right)^2} &= \frac{\max\left\{n\left(1-\lambda_n^{LR}\right)^2\left(\kappa_n^{LR}\right)^4, n\left(1-\lambda_n^{LR}\right)^2\left(\delta_n^{LR}\right)^8, n\left(1-\lambda_n^{LR}\right)^2\left(\kappa_n^{LR}\delta_n^{LR}\right)^2\right\}}{\left(w_{2n}^{LR}\right)^2} \\ &\leq \frac{\max\left\{\left(288 w_{2n}^{LR}\right)^2, \left(4 w_{1n}^{LR}\right)^2\right\}}{\left(w_{2n}^{LR}\right)^2} \leq 12^4 \times 4, \qquad \text{(A33)} \end{aligned}$$

where the first inequality in (A33) follows from the definition of D_n^c and the second one from the definition of F_n^c. But then,

$$\begin{aligned} \frac{LR_n(\theta_n^{LR})}{\left(w_{2n}^{LR}\right)^2} &= 2\frac{H_{3,n}}{\sqrt{n}} \frac{w_{1n}^{LR}}{\left(w_{2n}^{LR}\right)^2} + 2\frac{H_{4,n}}{\sqrt{n}} \frac{1}{w_{2n}^{LR}} - V_3 \frac{\left(w_{1n}^{LR}\right)^2}{\left(w_{2n}^{LR}\right)^2} - V_4 + \frac{r_n(\theta_n^{LR})}{\left(w_2^{LR}\right)^2} \\ &\leq 2\frac{M_1}{M} + 2\frac{M_1}{M} - V_4 + \left|\frac{r_n(\theta_n^{LR})}{h_n(\theta_n^{LR})}\right| \times 12^4 \times 4 \qquad \text{(A34)} \\ &\leq 4\xi - V_4 + \xi < 0, \qquad \text{(A35)} \end{aligned}$$

where (A34) follows from the combination of A_n with B_n, D_n^c, F_n^c and (A33), and (A35) follows from the definition of C_n and $V_4 = 24$.

To summarize, we have $A_n \cap B_n \cap C_n = \emptyset$, which contradicts

$$\Pr(A_n \cap B_n \cap C_n) \geq \frac{\epsilon}{2} \quad \text{i.o.},$$

as desired, and thus, $n(1-\lambda_n^{LR})^2(\delta_n^{LR})^8 = O_p(1)$ and $n(1-\lambda_n^{LR})^2(\kappa_n^{LR})^4 = O_p(1)$.

Step 3.2.2

Next, we will show that $n(1-\lambda_n^{LR})^2(\delta_n^{LR}\kappa_n^{LR})^2 = O_p(1)$, i.e. that for all $\epsilon > 0$, there exists $M > 1$ such that $\Pr[n(1-\lambda_n^{LR})^2\delta_n^{LR}\kappa_n^{LR2} > M] < \epsilon$ ult. To do so, notice that

$$r_n(\theta_n^{LR}) = o_p[h_n(\theta_n^{LR})] = o_p[\max\{1, n\left(1-\lambda_n^{LR}\right)^2\left(\delta_n^{LR}\kappa_n^{LR}\right)^2\}]$$

because $n(1-\lambda_n^{LR})^2(\delta_n^{LR})^8 = O_p(1)$ and $n(1-\lambda_n^{LR})^2(\kappa_n^{LR})^4 = O_p(1)$. Letting $0 < m < \frac{1}{4}V_3$, we have that

$$\Pr\left(\left|\frac{16 r_n(\theta_n^{LR})}{\max\{1, n\left(1-\lambda_n^{LR}\right)^2\left(\delta_n^{LR}\kappa_n^{LR}\right)^2\}}\right| > 2m\right) < \frac{\epsilon}{2} \text{ ult.} \tag{A36}$$

In turn, given that $H_{3,n}/\sqrt{n}$ and $H_{4,n}/\sqrt{n}$ are $O_p(1)$, there exists $M > 1$ such that for all n,

$$\Pr\left[\frac{H_{3,n}}{\sqrt{n}} \geq M\left(\frac{V_3}{2} - 2m\right)\right] < \frac{\epsilon}{4} \text{ and } \Pr\left[\frac{1}{2V_4}\left(\frac{H_{4,n}}{\sqrt{n}}\right)^2 > mM^2\right] < \frac{\epsilon}{4}. \tag{A37}$$

We then have that $\Pr\left(\left|w_{1n}^{LR}\right| > M\right)$ is equal to

$$\begin{aligned}
&= \Pr\left[\{\left|w_{1n}^{LR}\right| > M\} \cap \{LR(\theta_n^{LR}) \geq 0\}\right] \\
&= \Pr\left[\{\left|w_{1n}^{LR}\right| > M\} \cap \left\{\frac{LM_n^a(\theta_n^{LR})}{(w_{1n}^{LR})^2} + \frac{r_n(\theta_n^{LR})}{(w_{1n}^{LR})^2} \geq 0\right\}\right] \\
&= \Pr\left[\{\left|w_{1n}^{LR}\right| > M\} \cap \left\{\frac{LM_n^a(\theta_n^{LR})}{(w_{1n}^{LR})^2} + \frac{r_n(\theta_n^{LR})}{(w_{1n}^{LR})^2} \geq 0\right\} \cap \left\{\left|\frac{r_n(\theta_n^{LR})}{(w_{1n}^{LR})^2}\right| \leq 2m\right\}\right] \\
&\quad + \Pr\left[\{\left|w_{1n}^{LR}\right| > M\} \cap \left\{\frac{LM_n^a(\theta_n^{LR})}{(w_{1n}^{LR})^2} + \frac{r_n(\theta_n^{LR})}{(w_{1n}^{LR})^2} \geq 0\right\} \cap \left\{\left|\frac{r_n(\theta_n^{LR})}{(w_{1n}^{LR})^2}\right| > 2m\right\}\right] \\
&\leq \Pr\left[\{\left|w_{1n}^{LR}\right| > M\} \cap \left\{\frac{H_{3,n}}{\sqrt{n}}\frac{1}{w_{1n}^{LR}} - \frac{V_3}{2} - \frac{V_4\left(w_{2n}^{LR} - \frac{1}{V_4}\frac{H_{4,n}}{\sqrt{n}}\right)^2}{2(w_{1n}^{LR})^2} + \frac{\frac{1}{V_4}\left(\frac{H_{4,n}}{\sqrt{n}}\right)^2}{2(w_{1n}^{LR})^2} + m \geq 0\right\}\right] \\
&\quad + \Pr\left[\left|\frac{r_n(\theta_n^{LR})}{\left(w_{1n}^{LR}\right)^2}\right| > 2m\right] \\
&\leq \Pr\left(\{\left|w_{1n}^{LR}\right| > M\} \cap \left\{\frac{H_{3,n}}{\sqrt{n}} \geq w_{1n}^{LR}\left[\frac{V_3}{2} - m - \frac{H_{4,n}^2}{2nV_4}\frac{1}{\left(w_{1n}^{LR}\right)^2}\right]\right\}\right) + \frac{\epsilon}{2} && \text{(A38)} \\
&\leq \Pr\left[\frac{H_{3,n}}{\sqrt{n}} \geq M\left(\frac{V_3}{2} - m - \frac{H_{4,n}^2}{2nV_4}\frac{1}{M^2}\right)\right] + \frac{\epsilon}{2} \text{ ult.,} && \text{(A39)}
\end{aligned}$$

where (A38) uses (A36). In addition,

$$\begin{aligned}(A39) \leq& \Pr\left[\left\{\frac{H_{3,n}}{\sqrt{n}} \geq M\left(\frac{V_3}{2} - m - \frac{H_{4,n}^2}{2nV_4}\frac{1}{M^2}\right)\right\} \cap \left\{\frac{H_{4,n}^2}{2nV_4} \leq mM^2\right\}\right] \\ &+ \Pr\left[\left\{\frac{H_{3,n}}{\sqrt{n}} \geq M\left(\frac{V_3}{2} - m - \frac{H_{4,n}^2}{2nV_4}\frac{1}{M^2}\right)\right\} \cap \left\{\frac{H_{4,n}^2}{2nV_4} > mM^2\right\}\right] + \frac{\epsilon}{2} \\ \leq& \Pr\left[\frac{H_{3,n}}{\sqrt{n}} \geq M\left(\frac{V_3}{2} - 2m\right)\right] + \Pr\left(\frac{H_{4,n}^2}{2nV_4} > mM^2\right) + \frac{\epsilon}{2} \\ \leq& \frac{\epsilon}{4} + \frac{\epsilon}{4} + \frac{\epsilon}{2} = \epsilon, \qquad \text{(A40)}\end{aligned}$$

where in (A40) we have used (A37).

Step 4

We now show that $LR_n(\theta_n^{LR}) = LM_n^a(\theta_n^{LM}) + o_p(1)$, that is, that for all $\epsilon_1 > 0$ and for all $\epsilon_2 > 0$, there exists N such that for all $n > N$,

$$P\left(\left|LR_n(\theta_n^{LR}) - LM_n^a(\theta_n^{LM})\right| < \epsilon_1\right) > 1 - \epsilon_2.$$

Letting

$$\begin{aligned} G_n \;=\; & \Big\{n^{\frac{1}{2}}\left(1-\lambda_n^{LR}\right)\left(\delta_n^{LR}\right)^4, |n^{\frac{1}{2}}\left(1-\lambda_n^{LR}\right)\delta_n^{LR}\kappa_n^{LR}|, n^{\frac{1}{2}}\left(1-\lambda_n^{LR}\right)\left(\kappa_n^{LR}\right)^2, \\ & n^{\frac{1}{2}}\left(1-\lambda_n^{LR}\right)\left(\delta_n^{LM}\right)^4, |n^{\frac{1}{2}}\left(1-\lambda_n^{LR}\right)\delta_n^{LM}\kappa_n^{LM}|, n^{\frac{1}{2}}\left(1-\lambda_n^{LR}\right)\left(\kappa_n^{LM}\right)^2\Big\},\end{aligned}$$

we know that $\max\{G_n\} = O_p(1)$, so that for $\epsilon_2 > 0$ there exists M such that for all n,

$$\Pr\left(\max G_n \leq M\right) > 1 - \frac{\epsilon_2}{2}. \qquad \text{(A41)}$$

Letting $A = \{\theta \in \Theta : n^{\frac{1}{2}}(1-\lambda)\delta^4 \leq M, n^{\frac{1}{2}}(1-\lambda)\kappa^2 \leq M, |n^{\frac{1}{2}}(1-\lambda)\delta\kappa| \leq M\}$, we can then show

$$\sup_{\theta\in A}|LR_n(\theta) - LM_n^a(\theta)| = o_p(1),$$

i.e. there exists N such that for all $n > N$, we have that

$$\Pr\left(\sup_{\theta\in A}|LR_n(\theta) - LM_n^a(\theta)| < \epsilon_1\right) > 1 - \frac{\epsilon_2}{2}. \qquad \text{(A42)}$$

To show this, let

$$(\delta_n, \kappa_n, \lambda_n) \in \arg\max_{(\delta,\kappa,\lambda)\in A}\left|LR_n(\delta,\kappa,\lambda) - LM_n^a(\delta,\kappa,\lambda)\right|.$$

Given that $n^{\frac{1}{2}}(1-\lambda_n)\delta_n^4 = O_p(1)$ and $n^{\frac{1}{2}}(1-\lambda_n)\kappa_n^2 = O_p(1)$, we have $\delta_n, \kappa_n \xrightarrow{p} 0$, whence

$$\sup_{(\delta,\kappa,\lambda)\in A}|LR_n(\delta,\kappa,\lambda) - LM_n^a(\delta,\kappa,\lambda)| = |LR_n(\delta_n,\kappa_n,\lambda_n) - LM_n^a(\delta_n,\kappa_n,\lambda_n)| = o_p(1),$$

where the second equality follows from (A1). Therefore, for $n > N$ we have

$$\begin{aligned}
&\Pr\left(\left|LR_n(\theta_n^{LR}) - LM_n^a(\theta_n^{LM})\right| < \epsilon_1\right) \\
\geq &\Pr\left(\left\{\left|LR_n(\theta_n^{LR}) - LM_n^a(\theta_n^{LM})\right| < \epsilon_1\right\} \cap \left\{\theta_n^{LR} \in A\right\} \cap \left\{\theta_n^{LM} \in A\right\}\right) \\
\geq &\Pr\left(\left\{\sup_{\theta\in A}\left|LR_n(\theta) - LM_n^a(\theta)\right| < \epsilon_1\right\} \cap \left\{\theta_n^{LR} \in A\right\} \cap \left\{\theta_n^{LM} \in A\right\}\right) \qquad \text{(A43)} \\
\geq &\Pr\left(\sup_{\theta\in A}\left|LR_n(\theta) - LM_n^a(\theta)\right| < \epsilon_1\right) + P\left(\left\{\theta_n^{LR} \in A\right\} \cap \left\{\theta_n^{LM} \in A\right\}\right) - 1 \qquad \text{(A44)} \\
\geq &1 - \frac{\epsilon_2}{2} + 1 - \frac{\epsilon_2}{2} - 1 = 1 - \epsilon_2, \qquad \text{(A45)}
\end{aligned}$$

where we have used $\Pr(E_1 \cap E_2) \geq \Pr(E_1) + \Pr(E_2) - 1$ to go from (A43) to (A44), and (A41) and (A42) to go from (A44) to (A45).

Step 5

We consider the different cases separately in Step 5.1: $\mathcal{P} = \mathcal{P}_{a,1}$, Step 5.2: $\mathcal{P} = \mathcal{P}_{a,2}$ and Step 5.3: $\mathcal{P} = \mathcal{P}_{a,3}$.

Step 5.1 We have that

$$LM_n^a(\delta, \kappa, \lambda) = -V_3\left(w_{1n} - \frac{1}{V_3}\frac{H_{3,n}}{\sqrt{n}}\right)^2 + \frac{1}{V_3}\left(\frac{H_{3,n}}{\sqrt{n}}\right)^2 - V_4\left(w_{2n} - \frac{1}{V_4}\frac{H_{4,n}}{\sqrt{n}}\right)^2 + \frac{1}{V_4}\left(\frac{H_{4,n}}{\sqrt{n}}\right)^2,$$

where

$$w_1 = -\frac{1}{2}(1-\lambda)\lambda\sqrt{n}\delta\kappa \quad \text{and} \quad w_2 = \lambda(1-\lambda)\sqrt{n}\left(\frac{1}{8}\kappa^2 - \frac{1-\lambda+\lambda^2}{36}\delta^4\right).$$

Next, let

$$w_{21} = \frac{(1-\lambda)\lambda}{8}\sqrt{n}\kappa^2 \quad \text{and} \quad w_{22} = -\frac{(1-\lambda)\lambda(1-\lambda+\lambda^2)}{36}\sqrt{n}\delta^4.$$

We first aim to find an upper bound for $LM_n^a(\theta_n^{LM})$. In that respect, we can easily show that

$$LM_n^a(\theta_n^{LM}) \leq \frac{H_{3,n}^2}{nV_3} + \frac{H_{4,n}^2}{nV_4}. \qquad \text{(A46)}$$

Second, we aim to find a lower bound for $LM_n^a(\theta_n^{LM})$. To do so, let $\lambda_n^* = 1/2$,

$$\delta_n^* = \begin{cases} 2n^{-\frac{1}{8}}\left(-\frac{12}{V_4}\frac{H_{4,n}}{\sqrt{n}}\right)^{\frac{1}{4}} & \text{if } H_{4,n} \leq 0, \\ -n^{-\frac{1}{4}}\left|\frac{2}{V_3}\frac{H_{3,n}}{\sqrt{n}}\right| \Big/ \sqrt{\frac{2}{V_4}\frac{H_{4,n}}{\sqrt{n}}} & \text{if } H_{4,n} > 0, \end{cases}$$

and

$$\kappa_n^* = \begin{cases} -\left(n^{-\frac{3}{8}}\frac{4}{V_3}\frac{H_{3,n}}{\sqrt{n}}\right) \Big/ \left(-\frac{12}{V_4}\frac{H_{4,n}}{\sqrt{n}}\right)^{\frac{1}{4}} & \text{if } H_{4,n} < 0, \\ 4\text{sign}(H_{3,n})n^{-\frac{1}{4}}\sqrt{\frac{2}{V_4}\frac{H_{4,n}}{\sqrt{n}}} & \text{if } H_{4,n} \geq 0. \end{cases}$$

It is then easy to verify that $(\delta_n^*, \kappa_n^*, \lambda_n^*) \in \mathcal{P}_a$ with probability approaching one, whence

$$LM_n^a\left(\theta_n^{LM}\right) \geq LM_n^a\left(\delta_n^*, \kappa_n^*, \lambda_n^*\right) = \frac{H_{3,n}^2}{nV_3} + \frac{H_{4,n}^2}{nV_4} + o_p(1). \qquad \text{(A47)}$$

To verify the second equality of (A47), we can easily check by hand that

$$w_1^* = -\frac{1}{2}(1-\lambda_n^*)\lambda_n^*\sqrt{n}\delta_n^*\kappa_n^* = \frac{1}{V_3}\frac{H_{3,n}}{\sqrt{n}},$$

$$w_{21}^* = \frac{(1-\lambda_n^*)\lambda_n^*}{8}\sqrt{n}(\kappa^*)^2 = \begin{cases} \frac{1}{32}n^{-\frac{1}{4}}\left(\frac{4}{V_3}\frac{H_{3,n}}{\sqrt{n}}\right)^2 \Big/ \left(-\frac{12}{V_4}\frac{H_{4,n}}{\sqrt{n}}\right)^{\frac{1}{2}} = o_p(1) & \text{if } H_{4,n} < 0, \\ \frac{1}{V_4}\frac{H_{4,n}}{\sqrt{n}} & \text{if } H_{4,n} \geq 0, \end{cases}$$

and

$$\begin{aligned} w_{22}^* &= -\frac{(1-\lambda_n^*)\lambda_n^*[1-\lambda_n^*+(\lambda_n^*)^2]}{36}\sqrt{n}(\delta_n^*)^4 \\ &= \begin{cases} \frac{1}{V_4}\frac{H_{4,n}}{\sqrt{n}} & \text{if } H_{4,n} \leq 0, \\ -\frac{1}{192}n^{-\frac{1}{2}}\left(-\left|\frac{2}{V_3}\frac{H_{3,n}}{\sqrt{n}}\right| \Big/ \sqrt{\frac{2}{V_4}\frac{H_{4,n}}{\sqrt{n}}}\right)^4 = o_p(1) & \text{if } H_{4,n} > 0, \end{cases} \end{aligned}$$

with

$$w_2^* = w_{21}^* + w_{22}^* = \frac{1}{V_4}\frac{H_{4,n}}{\sqrt{n}} + o_p(1).$$

But then, (A46) and (A47) imply that

$$LM_n^a\left(\theta_n^{LM}\right) = \frac{H_{3,n}^2}{nV_3} + \frac{H_{4,n}^2}{nV_4} + o_p(1).$$

Step 5.2: Recall that $\Theta_2 = \{\theta : \lambda \in [1/2, 1], \delta \in [-\underline{\delta}, \bar{\delta}], \kappa = (2\lambda-1)\delta^2/3]\}$. Then, given that $\kappa = (2\lambda-1)\delta^2/3$, we will have

$$w_1 = -\frac{(1-\lambda)\lambda(2\lambda-1)}{6}\sqrt{n}\delta^3 \quad \text{and} \quad w_2 = \frac{(1-\lambda)\lambda}{72}\left(-1-2\lambda+2\lambda^2\right)\sqrt{n}\delta^4.$$

As before, we first aim to find an upper bound for $LM_n^a(\theta_n^{LM})$. In that regard, we can notice that $w_2 \leq 0$ for $\theta \in \Theta_2$ so that

$$\begin{aligned} LM_n^a(\delta_n^{LM}, \kappa_n^{LM}, \lambda_n^{LM}) &\leq \frac{1}{V_3}\left(\frac{H_{3,n}}{\sqrt{n}}\right)^2 + \sup_{w_2 \in R^-}\left[-V_4\left(w_2 - \frac{1}{V_4}\frac{H_{4,n}}{\sqrt{n}}\right)^2 + \frac{1}{V_4}\left(\frac{H_{4,n}}{\sqrt{n}}\right)^2\right] \\ &= \frac{1}{V_3}\left(\frac{H_{3,n}}{\sqrt{n}}\right)^2 + \frac{1}{V_4}\left(\frac{H_{4,n}}{\sqrt{n}}\right)^2 \mathbf{1}\left[H_{4,n} < 0\right]. \end{aligned}$$

Second, we aim to find a lower bound for $LM_n^a(\theta_n^{LM})$. For that purpose, let $\bar{\lambda} \in (1/2, 1)$,

$$\delta_n^* = \begin{cases} -\text{sign}(H_{3,n})2n^{-\frac{1}{8}}\left(-\frac{12}{V_4}\frac{H_{4,n}}{\sqrt{n}}\right)^{\frac{1}{4}} & \text{if } H_{4,n} < 0, \\ -n^{-\frac{1}{6}}\left(\frac{\frac{6}{V_3}\frac{H_{3,n}}{\sqrt{n}}}{(1-\bar{\lambda})\bar{\lambda}(2\bar{\lambda}-1)}\right)^{\frac{1}{3}} & \text{if } H_{4,n} \geq 0, \end{cases}$$

and

$$\lambda_n^* = \begin{cases} \frac{1}{2} + n^{-\frac{1}{8}}\frac{\text{sign}(H_{3,n})\frac{3}{V_3}\frac{H_{3,n}}{\sqrt{n}}}{2\left(-\frac{12}{V_4}\frac{H_{4,n}}{\sqrt{n}}\right)^{\frac{3}{4}}} & \text{if } H_{4,n} < 0, \\ \bar{\lambda} & \text{if } H_{4,n} \geq 0. \end{cases}$$

We can then verify that

$$w_1^* = -\frac{(1-\lambda_n^*)\lambda_n^*(2\lambda_n^*-1)}{6}\sqrt{n}(\delta_n^*)^3 = \frac{1}{V_3}\frac{H_{3,n}}{\sqrt{n}} + o_p(1),$$

$$\begin{aligned} w_2^* &= \frac{(1-\lambda_n^*)\lambda_n^*}{72}[-1-2\lambda_n^* + 2(\lambda_n^*)^2]\sqrt{n}(\delta_n^*)^4 \\ &= \begin{cases} \frac{1}{V_4}\frac{H_{4,n}}{\sqrt{n}} + o_p(1) & \text{if } H_{4,n} < 0, \\ \frac{(1-\bar{\lambda})\bar{\lambda}}{72}(-1-2\bar{\lambda}+2\bar{\lambda}^2)n^{-\frac{1}{6}}\left[\frac{1}{(1-\bar{\lambda})\bar{\lambda}(2\bar{\lambda}-1)}\frac{6}{V_3}\frac{H_{3,,n}}{\sqrt{n}}\right]^{\frac{4}{3}} = o_p(1) & \text{if } H_{4,n} \geq 0. \end{cases} \end{aligned}$$

As a result,

$$LM_n^a(\theta_n^{LM}) \geq LM_n^a(\delta_n^*, \kappa_n^*, \lambda_n^*) = \frac{H_{3,n}^2}{nV_3} + \frac{H_{4,n}^2}{nV_4}\mathbf{1}[H_{4,n} < 0] + o_p(1),$$

whence

$$LM_n^a(\theta_n^{LM}) = \frac{H_{3,n}^2}{nV_3} + \frac{H_{4,n}^2}{nV_4}\mathbf{1}[H_{4,n} < 0],$$

as desired.

Step 5.3: Recall that $\Theta_3' = \{\vartheta : \lambda \in [1/2, 1], \delta = 0, \varkappa \in [-\underline{\kappa}, \bar{\kappa}]\}$ and $\mathcal{P}_{a,3} = \{(\delta, \kappa, \lambda) : (\delta, \kappa - (2\lambda-1)\delta^3/3, \lambda) \in \Theta_3', \max\{|\delta|, |\kappa|\} \leq 1-\lambda\}$. Exploiting the fact that $\delta = 0$, we have

$$w_1 = 0 \quad \text{and} \quad w_2 = \frac{1}{8}\lambda(1-\lambda)\sqrt{n}\kappa^2.$$

Thus,

$$LM_n^a(\delta, \kappa, \lambda) = -V_4\left(w_2 - \frac{1}{V_4}\frac{H_{4,n}}{\sqrt{n}}\right)^2 + \frac{1}{V_4}\left(\frac{H_{4,n}}{\sqrt{n}}\right)^2.$$

Next, we first aim to find an upper bound for $LM_n^a(\theta_n^{LM})$. It is easy to see that $w_2 \geq 0$ for $\theta \in \Theta_3$ so that

$$\begin{aligned} LM_n^a(\delta_n^{LM}, \kappa_n^{LM}, \lambda_n^{LM}) &\leq \sup_{w_2 \in R^+}\left[-V_4\left(w_2 - \frac{1}{V_4}\frac{H_{4,n}}{\sqrt{n}}\right)^2 + \frac{1}{V_4}\left(\frac{H_{4,n}}{\sqrt{n}}\right)^2\right] \\ &= \frac{1}{V_4}\left(\frac{H_{4,n}}{\sqrt{n}}\right)^2\mathbf{1}\left[H_{4,n} > 0\right]. \end{aligned}$$

Second, to find a lower bound for $LM_n^a(\theta_n^{LM})$, let $\lambda_n^* = 1/2$ and

$$\kappa_n^* = \begin{cases} 0 & \text{if } H_{4,n} \leq 0, \\ 4n^{-\frac{1}{4}}\sqrt{\frac{2H_{4,n}}{V_4\sqrt{n}}} & \text{if } H_{4,n} > 0. \end{cases}$$

As a result, $w_2^* = \frac{1}{V_4}\frac{H_{4,n}}{\sqrt{n}}\mathbf{1}\left[H_{4,n} > 0\right]$, whence

$$LM_n^a(\theta_n^{LM}) \geq LM_n^a(0, \kappa_n^*, \lambda_n^*) = \frac{H_{4,n}^2}{nV_4}\mathbf{1}[H_{4,n} \geq 0],$$

as desired. □

B Detailed proof of Proposition 3

Before proceeding with the proof, we start by giving an example of sequences $(\delta_m, \kappa_{1m}) \to 0$ and $(\delta_{2m}, \kappa_{2m}) \to 0$ such that

$$\lim_{m\to\infty} \frac{\mathcal{G}_n(\delta_{1m}, \kappa_{1m})}{\sqrt{V(\delta_{1m}, \kappa_{1m})}} \neq \lim_{m\to\infty} \frac{\mathcal{G}_n(\delta_{2m}, \kappa_{2m})}{\sqrt{V(\delta_{2m}, \kappa_{2m})}}.$$

Note that for $(\delta, \kappa) \to (0, 0)$, it holds

$$\frac{1}{\sqrt{n}} \frac{\partial L}{\partial \lambda}\bigg|_{(\kappa,\delta,1)} = \frac{H_{4n}}{\sqrt{n}} \left(\frac{1}{36}\delta^4 - \frac{1}{8}\kappa^2\right) + \frac{H_{3n}}{\sqrt{n}} \frac{1}{2}\delta\kappa + o_p[\tau(\kappa, \delta)],$$

where

$$\tau(\kappa, \delta) = \max\left\{\left|\frac{1}{36}\delta^4 - \frac{1}{8}\kappa^2\right|, \left|\frac{1}{2}\delta\kappa\right|\right\}.$$

Let

$$(\delta_1, \kappa_1) = \left(\sqrt{3\upsilon}, \sqrt{2}\upsilon\right), \qquad (\delta_2, \kappa_2) = (0, \upsilon).$$

It is easy to see that with $\upsilon \to 0$, we have $|\delta_1 - \delta_2| + |\kappa_1 - \kappa_2| \to 0$,

$$\lim_{\upsilon\to 0} \frac{\frac{1}{\sqrt{n}} \frac{\partial L}{\partial \lambda}(\delta_1, \kappa_1, 1)}{\sqrt{\operatorname{var}\left(\frac{\partial \ell}{\partial \lambda}(\delta_1, \kappa_1, 1)\right)}} = \frac{H_{3n}}{\sqrt{nV_3}} \quad \text{and} \quad \lim_{\upsilon\to 0} \frac{\frac{1}{\sqrt{n}} \frac{\partial L}{\partial \lambda}(\delta_2, \kappa_2, 1)}{\sqrt{\operatorname{var}\left(\frac{\partial \ell}{\partial \lambda}(\delta_2, \kappa_2, 1)\right)}} = \frac{H_{4n}}{\sqrt{nV_4}}.$$

This shows that the process $\frac{\mathcal{G}_n(\delta,\kappa)}{\sqrt{V(\delta,\kappa)}}$ is not stochastically equicontinuous.

Next, we follow the steps of the proof outlined in the appendix of the paper.

Step 1

Lemma 1 *Let $R_n^d(\eta, \tau, \varphi) = LR_n^d(\eta, \tau, \varphi) - LM_n^d(\eta, \tau, \varphi)$. For all sequences of $(\eta_n, \tau_n, \varphi_n) \in D^1$ and $\eta_n \xrightarrow{p} 0$, we have that*

$$R_n^d(\eta_n, \tau_n, \varphi_n) = o_p\left(\max\left\{1, n\eta_n^2\right\}\right).$$

Proof. Let $\delta_n = \delta(\tau_n, \varphi_n)$, $\kappa_n = \kappa(\tau_n, \varphi_n)$, $\lambda_n = \lambda(\eta_n, \tau_n, \varphi_n)$. First we show that $1 - \lambda_n \xrightarrow{p} 0$. Recall that $\eta_n = \max\left\{\left|\frac{1}{36}\delta_n^4 - \frac{1}{8}\kappa_n^2\right|, \left|\frac{1}{2}\delta_n\kappa_n\right|\right\}(1 - \lambda_n)$, whence either $(1 - \lambda_n) \leq \sqrt{\eta_n}$ or

$$\max\left\{\left|\frac{1}{36}\delta_n^4 - \frac{1}{8}\kappa_n^2\right|, \left|\frac{1}{2}\delta_n\kappa_n\right|\right\} \leq \sqrt{\eta_n}. \tag{B1}$$

Under (B1), we have

$$2\eta_n \geq \left(\frac{1}{36}\delta_n^4 - \frac{1}{8}\kappa_n^2\right)^2 + \frac{1}{4}\delta_n^2\kappa_n^2 = \left(\frac{1}{36}\delta_n^4\right)^2 + \left(\frac{1}{8}\kappa_n^2\right)^2 + \frac{1}{4}\delta_n^2\kappa_n^2\left(1 - \frac{1}{36}\delta_n^2\right). \tag{B2}$$

It is then easy to verify that given (B1), $1 - \frac{1}{36}\delta_n^2 \geq 0$ with probability approaching 1. Therefore, (B2) implies that

$$
\begin{aligned}
2\eta_n \geq & \left(\frac{1}{36}\delta_n^4\right)^2 + \left(\frac{1}{8}\kappa_n^2\right)^2 \\
\Rightarrow & |\delta_n| \leq 2^{5/8}\sqrt{3}\eta_n^{1/8}, |\kappa_n| \leq 2^{7/4}\eta_n^{1/4},
\end{aligned}
$$

and also, that $1-\lambda_n \leq \max\{|\delta_n|, |\kappa_n|\} \leq \max\{2^{5/8}\sqrt{3}\eta_n^{1/8}, 2^{7/4}\eta_n^{1/4}\}$ because of the restriction on $\mathcal{P}_b$. In sum, it holds that

$$
1-\lambda_n \leq \max\{2^{5/8}\sqrt{3}\eta_n^{1/8}, 2^{7/4}\eta_n^{1/4}, \eta_n^{1/2}\} \xrightarrow{p} 0.
$$

Second, a third-order Taylor expansion gives

$$
\begin{aligned}
\frac{1}{2}LR_n^d(\eta_n, \tau_n, \varphi_n) &= L_n^d(\eta_n, \tau_n, \varphi_n) - L_n^d(0, \tau_n, \varphi_n) \\
&= L_n(\delta_n, \kappa_n, \lambda_n) - L_n(\delta_n, \kappa_n, 1) \\
&= \frac{\partial L_n(\delta_n, \kappa_n, 1)}{\partial \lambda}(\lambda_n - 1) + \frac{1}{2}\frac{\partial^2 L_n(\delta_n, \kappa_n, 1)}{\partial \lambda^2}(\lambda_n - 1)^2 \\
&+ \frac{1}{3!}\frac{\partial^3 L_n(\delta_n, \kappa_n, \tilde{\lambda}_n)}{\partial \lambda^3}(\lambda_n - 1)^3.
\end{aligned}
$$

The first term is

$$
\begin{aligned}
\frac{\partial L_n(\delta_n, \kappa_n, 1)}{\partial \lambda}(\lambda_n - 1) &= \frac{1}{\sqrt{n}}\frac{1}{\tau_n}\frac{\partial L_n(\delta_n, \kappa_n, 1)}{\partial \lambda}\sqrt{n}\tau_n(\lambda_n - 1) \\
&= \mathcal{G}_n^d(\tau_n, \varphi_n)\sqrt{n}\tau_n(\lambda_n - 1).
\end{aligned}
$$

In turn, the second term will be

$$
\begin{aligned}
\frac{1}{2}\left\{\frac{1}{n}\frac{\partial^2 L_n(\delta_n, \kappa_n, 1)}{\partial \lambda^2}\right\}n(\lambda_n - 1)^2 &= \frac{1}{2}\left\{E\left[\frac{\partial^2 l(\delta_n, \kappa_n, 1)}{\partial \lambda^2}\right] + O_p\left(\frac{\tau_n}{\sqrt{n}}\right)\right\}n(\lambda_n - 1)^2 && \text{(B3)} \\
&= \frac{1}{2}E\left[\frac{\partial^2 l(\delta_n, \kappa_n, 1)}{\partial \lambda^2}\right]n(\lambda_n - 1)^2 + O_p[\sqrt{n}\tau_n(\lambda_n - 1)^2] \\
&= \frac{1}{2}E\left[\tau_n^{-2}\frac{\partial^2 l(\delta_n, \kappa_n, 1)}{\partial \lambda^2}\right]n\tau_n^2(\lambda_n - 1)^2 + O_p[\sqrt{n}\tau_n(\lambda_n - 1)^2] && \text{(B4)} \\
&= -\frac{1}{2}V^d(\tau_n, \varphi_n)n\tau_n^2(\lambda_n - 1)^2 + o_p[\sqrt{n}\tau_n(\lambda_n - 1)], && \text{(B5)}
\end{aligned}
$$

where (B3) follows from Lemma 8(8.1), and (B4) to (B5) from the information matrix equality.

Let us now turn to the third term. In view of Lemmas 8.2 and 8.5, we have

$$
\begin{aligned}
\left|\frac{1}{n}\tau_n^{-1}\frac{\partial^3 L(\delta_n, \kappa_n, \tilde{\lambda}_n)}{\partial \lambda^3}\right| &= \left|\tau_n^{-1}E\left[\frac{\partial^3 l(\delta_n, \kappa_n, \tilde{\lambda}_n)}{\partial \lambda^3}\right] + O_p\left(\frac{1}{\sqrt{n}}\right)\right| \\
&= O(\tau_n) + O_p\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}
$$

whence

$$\frac{1}{n}\frac{\partial^3 L(\delta_n,\kappa_n,\tilde{\lambda}_n)}{\partial\lambda^3}n(\lambda_n-1)^3=\left[O(\tau_n)+O_p\left(\frac{1}{\sqrt{n}}\right)\right]n\tau_n(\lambda_n-1)^3=o_p[n\tau_n^2(\lambda_n-1)^2].$$

In sum, we have $LR(\delta_n,\kappa_n,\lambda_n)=LM(\delta_n,\kappa_n,\lambda_n)+o_p\left(n\eta_n^2\right)$. □

Step 2

Lemma 2 *For $(\tau,\varphi)\in D^1_{\tau\varphi}$, $\mathcal{G}^d_n(\tau,\varphi)\Rightarrow\mathcal{G}^d(\tau,\varphi)$, where $\mathcal{G}^d(\tau,\varphi)$ is a Gaussian process with mean 0 and covariance kernel*

$$\mathcal{K}[(\tau,\varphi),(\tau',\varphi')]=\frac{1}{\tau\tau'}cov\left\{\frac{\partial l[\delta(\tau,\varphi),\kappa(\tau,\varphi),1]}{\partial\lambda},\frac{\partial l[\delta(\tau',\varphi'),\kappa(\tau',\varphi'),1]}{\partial\lambda}\right\}. \tag{B6}$$

Proof. Here we follow Andrews (2001). By Theorem 10.2 of Pollard (1990), $\mathcal{G}^d_n(\cdot)\Rightarrow\mathcal{G}^d(\cdot)$ if (i) the domain of (τ,φ) is totally bounded, (ii) the finite dimensional distributions of $\mathcal{G}^d_n(\cdot)$ converge to those of $\mathcal{G}^d(\cdot)$, (iii) $\left\{\mathcal{G}^d_n(\cdot):n\geq 1\right\}$ is stochastically equicontinuous.

(i) is satisfied because $(\tau,\varphi)\subset[0,\bar{\delta}^4+\bar{\kappa}^2+\bar{\delta}\bar{\kappa}]\times[0,1]$.

(ii) The process $\tau^{-1}\partial l_i(\delta(\tau,\varphi),\kappa(\tau,\varphi),1)/\partial\lambda$ is *iid* with mean 0.

Moreover,

$$E\left[\sup_{(\tau,\varphi)\in D^1_{\tau\varphi}}\left|\frac{1}{\tau}\frac{\partial l(\delta(\tau,\varphi),\kappa(\tau,\varphi),1)}{\partial\lambda}\right|\right]\leq E\left[\sup_{|\delta|\leq\bar{\delta}^2,|\kappa|\leq\bar{\kappa}^2,\delta^2+\kappa^2>0}\left|\frac{1}{\tau(\delta,\kappa)}\frac{\partial l(\delta,\kappa,1)}{\partial\lambda}\right|\right]<\infty. \tag{B7}$$

To prove (B7), consider the fifth-order Taylor expansion of $\partial l\left(\delta,\kappa,1\right)/\partial\lambda$ around $(\delta,\kappa)=(0,0)$ given by

$$\begin{aligned}\frac{\partial l\left(\delta,\kappa,1\right)}{\partial\lambda}=&\sum_{k=1}^{4}\sum_{i+j=k}\frac{1}{i!j!}\frac{\partial^{1+k}l\left(0,0,1\right)}{\partial\lambda\partial\delta^i\partial\kappa^j}\delta^i\kappa^j+\sum_{i+j=5}\frac{1}{i!j!}\frac{\partial^6 l(\tilde{\delta},\tilde{\kappa},1)}{\partial\lambda\partial\delta^i\partial\kappa^j}\delta^i\kappa^j\\=&h_4\left(\frac{1}{36}\delta^4-\frac{1}{8}\kappa^2\right)+h_3\frac{1}{2}\delta\kappa+\sum_{4\geq i+j\geq 3,i\geq 1,j\geq 1}\frac{1}{i!j!}\frac{\partial^6 l\left(0,0,1\right)}{\partial\lambda\partial\delta^i\partial\kappa^j}\delta^i\kappa^j\\&+\sum_{i+j=5,i\geq 1,j\geq 1}\frac{1}{i!j!}\frac{\partial^6 l(\tilde{\delta},\tilde{\kappa},1)}{\partial\lambda\partial\delta^i\partial\kappa^j}\delta^i\kappa^j+\frac{\partial^6 l(\tilde{\delta},\tilde{\kappa},1)}{\partial\lambda\partial\delta^5}\delta^5\\&+\left[\frac{\partial^4 l\left(0,0,1\right)}{\partial\lambda\partial\kappa^3}+\frac{\partial^5 l\left(0,0,1\right)}{\partial\lambda\partial\kappa^4}\kappa+\frac{\partial^6 l(\tilde{\delta},\tilde{\kappa},1)}{\partial\lambda\partial\kappa^5}\kappa^2\right]\kappa^3.\end{aligned} \tag{B8}$$

Consequently

$$\begin{aligned}\left|\frac{1}{\tau(\delta,\kappa)}\frac{\partial L(\delta,\kappa,1)}{\partial\lambda}\right|\leq&|h_4|+|h_3|+\sum_{4\geq i+j\geq 3,i\geq 1,j\geq 1}\left|\frac{\partial^6 l\left(0,0,1\right)}{\partial\lambda\partial\delta^i\partial\kappa^j}\right|\frac{2}{i!j!}\bar{\delta}^{i-1}\bar{\kappa}^{j-1}\\+&\sum_{i+j=5,i\geq 1,j\geq 1}\sup_{|\tilde{\delta}|\leq\bar{\delta},|\tilde{\kappa}|\leq\bar{\kappa}}\left|\frac{1}{i!j!}\frac{\partial^6 l(\tilde{\delta},\tilde{\kappa},1)}{\partial\lambda\partial\delta^i\partial\kappa^j}\right|\bar{\delta}^{i-1}\bar{\kappa}^{j-1}+\sup_{|\tilde{\delta}|\leq\bar{\delta},|\tilde{\kappa}|\leq\bar{\kappa}}\left|\frac{\partial^6 l(\tilde{\delta},\tilde{\kappa},1)}{\partial\lambda\partial\delta^5}\right|\left|\frac{\delta^5}{\tau(\delta,\kappa)}\right|\\+&\left(\left|\frac{\partial^4 l\left(0,0,1\right)}{\partial\lambda\partial\kappa^3}\right|+\left|\frac{\partial^5 l\left(0,0,1\right)}{\partial\lambda\partial\kappa^4}\right||\bar{\kappa}|+\sup_{|\tilde{\delta}|\leq\bar{\delta},|\tilde{\kappa}|\leq\bar{\kappa}}\left|\frac{\partial^6 l(\tilde{\delta},\tilde{\kappa},1)}{\partial\lambda\partial\kappa^5}\right|\bar{\kappa}^2\right)\left|\frac{\kappa^3}{\tau(\delta,\kappa)}\right|.\end{aligned} \tag{B9}$$

It is then easy to check that

$$E\left[|h_4|+|h_3|+\sum_{4\geq i+j\geq 3, i\geq 1, j\geq 1}\frac{1}{i!j!}\left|\frac{\partial^6 l\left(0,0,1\right)}{\partial\lambda\partial\delta^i\partial\kappa^j}\right|+\left|\frac{\partial^4 l\left(0,0,1\right)}{\partial\lambda\partial\kappa^3}\right|+\left|\frac{\partial^5 l\left(0,0,1\right)}{\partial\lambda\partial\kappa^4}\right|\right]<\infty \quad \text{(B10)}$$

and

$$E\left[\sum_{i+j=5, i\geq 0, j\geq 0}\sup_{|\tilde{\delta}|\leq\bar{\delta}, |\tilde{\kappa}|\leq\bar{\kappa}}\left|\frac{\partial^6 l(\tilde{\delta},\tilde{\kappa},1)}{\partial\lambda\partial\delta^i\partial\kappa^j}\right|\right]<\infty. \quad \text{(B11)}$$

For $\delta^2+\kappa^2>0$, if $\kappa=0$, $\kappa^2/\max\{|\frac{1}{36}\delta^4-\frac{1}{8}\kappa^2|, |\frac{1}{2}\delta\kappa|\}=0$, otherwise

$$\frac{\kappa^2}{\tau(\delta,\kappa)}=\frac{1}{\max\left\{\left|\frac{1}{36}\frac{\delta^2}{\kappa^2}\delta^2-\frac{1}{8}\right|, \left|\frac{1}{2}\frac{\delta}{\kappa}\right|\right\}}\leq\begin{cases}\frac{2}{\left|\frac{\delta}{\kappa}\right|}\leq 2\bar{\delta} & \text{if } \delta^2/\kappa^2\geq\bar{\delta}^{-2}, \\ \frac{1}{\left|\frac{1}{36}\frac{\delta^2}{\kappa^2}\delta^2-\frac{1}{8}\right|}\leq\frac{1}{\left|\frac{1}{36}-\frac{1}{8}\right|}=\frac{72}{7} & \text{if } \delta^2/\kappa^2\leq\bar{\delta}^{-2}.\end{cases} \quad \text{(B12)}$$

Finally,

$$\left|\frac{\delta^5}{\tau}\right|\leq\delta\left(\frac{\left|\delta^4-\frac{36}{8}\kappa^2\right|}{\max\left\{\left|\frac{1}{36}\delta^4-\frac{1}{8}\kappa^2\right|, \left|\frac{1}{2}\delta\kappa\right|\right\}}+\frac{\frac{36}{8}\kappa^2}{\max\left\{\left|\frac{1}{36}\delta^4-\frac{1}{8}\kappa^2\right|, \left|\frac{1}{2}\delta\kappa\right|\right\}}\right)<36\bar{\delta}\left[1+\frac{1}{8}\left(2\bar{\delta}+\frac{72}{7}\right)\right] \quad \text{(B13)}$$

In sum, (B7) follows from (B9)–(B13). But given (B7), the martingale difference central limit theorem of Billingsley (1968, Theorem 3.1) implies that each of the finite dimensional distributions of $\mathcal{G}_n^d(\cdot)$ converges in distribution to a multivariate normal distribution with covariance given by (B6).

(iii) The process $\mathcal{G}_n^d(\tau,\varphi)$ is stochastically equicontinuous if for all $\varepsilon>0$, there exists $c>0$ such that

$$\limsup_{n\to\infty}\Pr\left[\sup_{\|(\tau_1,\varphi_1)-(\tau_2,\varphi_2)\|\leq c,\ (\tau_1,\varphi_1),(\tau_2,\varphi_2)\in D^1_{\tau\varphi}}\left|\mathcal{G}_n^d(\tau_1,\varphi_1)-\mathcal{G}_n^d(\tau_2,\varphi_2)\right|>\varepsilon\right]<\varepsilon. \quad \text{(B14)}$$

In the rest of this section, we keep the restriction $(\tau_1,\varphi_1),(\tau_2,\varphi_2)\in D^1_{\tau\varphi}$ implicit to simplify notation.

The proof has two steps. First, we show that for all $\varepsilon>0$, there exist $c_1\geq c_2>0$ such that

$$\Pr\left[\sup_{\|(\tau_1,\varphi_1)-(\tau_2,\varphi_2)\|\leq c_2, |\tau_1|,|\tau_2|\leq 2c_1}\left|\mathcal{G}_n^d(\tau_1,\varphi_1)-\mathcal{G}_n^d(\tau_2,\varphi_2)\right|>\varepsilon\right]\leq\frac{\varepsilon}{2}. \quad \text{(B15)}$$

Second, we show that given c_1 above, there is $c_1\geq c_3>0$ such that

$$\Pr\left[\sup_{\|(\tau_1,\varphi_1)-(\tau_2,\varphi_2)\|\leq c_3, |\tau_1|,|\tau_2|\geq c_1}\left|\mathcal{G}_n^d(\tau_1,\varphi_1)-\mathcal{G}_n^d(\tau_2,\varphi_2)\right|>\varepsilon\right]\leq\frac{\varepsilon}{2}. \quad \text{(B16)}$$

Let $c = \min\{c_2, c_3\}$. Whence (B14) follows from

$$
\begin{aligned}
&\Pr\left[\sup_{\|(\tau_1,\varphi_1)-(\tau_2,\varphi_2)\|\le c}\left|\mathcal{G}_n^d(\tau_1,\varphi_1)-\mathcal{G}_n^d(\tau_2,\varphi_2)\right|>\varepsilon\right] \\
\le &\Pr\left[\left\{\sup_{\|(\tau_1,\varphi_1)-(\tau_2,\varphi_2)\|\le c,|\tau_1|,|\tau_2|\le 2c_1}\left|\mathcal{G}_n^d(\tau_1,\varphi_1)-\mathcal{G}_n^d(\tau_2,\varphi_2)\right|>\varepsilon\right\}\right. \\
&\quad \left.\cup\left\{\sup_{\|(\tau_1,\varphi_1)-(\tau_2,\varphi_2)\|\le c,|\tau_1|,|\tau_2|\ge c_1}\left|\mathcal{G}_n^d(\tau_1,\varphi_1)-\mathcal{G}_n^d(\tau_2,\varphi_2)\right|>\varepsilon\right\}\right] \\
\le &\Pr\left[\sup_{\|(\tau_1,\varphi_1)-(\tau_2,\varphi_2)\|\le c_2,|\tau_1|,|\tau_2|\le 2c_1}\left|\mathcal{G}_n^d(\tau_1,\varphi_1)-\mathcal{G}_n^d(\tau_2,\varphi_2)\right|>\varepsilon\right] \qquad \text{(B17)} \\
&\quad +\Pr\left[\sup_{\|(\tau_1,\varphi_1)-(\tau_2,\varphi_2)\|\le c_3,|\tau_1|,|\tau_2|\ge c_1}\left|\mathcal{G}_n^d(\tau_1,\varphi_1)-\mathcal{G}_n^d(\tau_2,\varphi_2)\right|>\varepsilon\right]\le\varepsilon \qquad \text{(B18)}
\end{aligned}
$$

where the first inequality follows from that for $0 < c \le c_1$,

$$
\begin{aligned}
&\sup_{\|(\tau_1,\varphi_1)-(\tau_2,\varphi_2)\|\le c}\left|\mathcal{G}_n^d(\tau_1,\varphi_1)-\mathcal{G}_n^d(\tau_2,\varphi_2)\right| \qquad \text{(B19)} \\
\le&\max\left\{\sup_{\|(\tau_1,\varphi_1)-(\tau_2,\varphi_2)\|\le c,|\tau_1|,|\tau_2|\le 2c_1}\left|\mathcal{G}_n^d(\tau_1,\varphi_1)-\mathcal{G}_n^d(\tau_2,\varphi_2)\right|,\right. \\
&\quad\left.\sup_{\|(\tau_1,\varphi_1)-(\tau_2,\varphi_2)\|\le c,|\tau_1|,|\tau_2|\ge c_1}\left|\mathcal{G}_n^d(\tau_1,\varphi_1)-\mathcal{G}_n^d(\tau_2,\varphi_2)\right|\right\},
\end{aligned}
$$

and the second inequality follows from $c \le c_2$ and $c \le c_3$.

We next show that there exist $c_1 \ge c_2 > 0$ such that (B15) holds. Given (B8), we will have that

$$
\begin{aligned}
\mathcal{G}_n^d(\tau,\varphi) =&\frac{H_4}{\sqrt{n}}\frac{\frac{1}{36}\delta^4-\frac{1}{8}\kappa^2}{\tau}+\frac{H_3}{\sqrt{n}}\frac{\frac{1}{2}\delta\kappa}{\tau}+\sum_{4\ge i+j\ge 3,j\ge 1}\frac{1}{i!j!}\frac{1}{\sqrt{n}}\frac{\partial^6 L_n(0,0,1)}{\partial\lambda\partial\delta^i\partial\kappa^j}\frac{\delta^i\kappa^j}{\tau} \\
&+\sum_{i+j=5}\frac{1}{i!j!}\frac{1}{\sqrt{n}}\frac{\partial^6 L_n\left(\tilde{\delta},\tilde{\kappa},1\right)}{\partial\lambda\partial\kappa^5}\frac{\delta^i\kappa^j}{\tau},
\end{aligned}
$$

where $|\tilde{\delta}| \le |\delta|$, $|\tilde{\kappa}| \le |\kappa|$, and δ, κ, $\tilde{\delta}$, $\tilde{\kappa}$ are functions of (τ, φ) even though we have omitted these arguments. Therefore,

$$
\begin{aligned}
&\frac{1}{21}\left[\mathcal{G}_n^d(\tau_1,\varphi_1)-\mathcal{G}_n^d(\tau_2,\varphi_2)\right]^2 \\
\leq &\left(\frac{H_4}{\sqrt{n}}\right)^2\left\{\tau_1^{-1}\left(\frac{1}{36}\delta_1^4-\frac{1}{8}\kappa_1^2\right)-\tau_2^{-1}\left(\frac{1}{36}\delta_2^4-\frac{1}{8}\kappa_2^2\right)\right\}^2 &\text{(B20)}\\
+&\left(\frac{H_3}{\sqrt{n}}\right)^2\left\{\frac{1}{2}\tau_1^{-1}\delta_1\kappa_1-\frac{1}{2}\tau_2^{-1}\delta_2\kappa_2\right\}^2 &\text{(B21)}\\
+&\sum_{4\geq i+j\geq 3, j\geq 1}\left(\frac{1}{i!j!}\frac{1}{\sqrt{n}}\frac{\partial^6 L_n(0,0,1)}{\partial\lambda\partial\delta^i\partial\kappa^j}\right)^2\left\{\tau_1^{-1}\delta_1^i\kappa_1^j-\tau_2^{-1}\delta_2^i\kappa_2^j\right\}^2 &\text{(B22)}\\
+&\sum_{i+j=5}\sup_{|\delta|\leq\bar{\delta},|\kappa|\leq\bar{\kappa}}\left(\frac{1}{i!j!}\frac{1}{\sqrt{n}}\frac{\partial^6 L_n(\delta,\kappa,1)}{\partial\lambda\partial\delta^i\partial\kappa^j}\right)^2\left\{\tau_1^{-2}\delta_1^{2i}\kappa_1^{2j}+\tau_2^{-2}\delta_2^{2i}\kappa_2^{2j}\right\}, &\text{(B23)}
\end{aligned}
$$

where $\delta_1=\delta(\tau_1,\varphi_1)$, $\kappa_1=\kappa(\tau_1,\varphi_1)$, δ_2 and κ_2 are defined in the same way. First, we can easily check that

$$E\left[\left(\frac{H_4}{\sqrt{n}}\right)^2\right]=E\left[h_4^2\right]<\infty,\ E\left[\left(\frac{H_3}{\sqrt{n}}\right)^2\right]=E\left[h_3^2\right]<\infty$$

and

$$E\left[\left(\frac{1}{i!j!}\frac{1}{\sqrt{n}}\frac{\partial^6 L_n(0,0,1)}{\partial\lambda\partial\delta^i\partial\kappa^j}\right)^2\right]=E\left(\frac{1}{i!j!}\frac{\partial^6 l(0,0,1)}{\partial\lambda\partial\delta^i\partial\kappa^j}\right)^2<\infty.$$

by the *iid* assumption and the zero expectation of these terms. Second, for the terms (B20)-(B23), we can show that the non-random coefficients in {} converge to zero as $c_1,c_2\to 0$, using arguments in (B12), (B13) and Lemma 9. To be more specific, for $(\tau,\varphi)\in D^1$, we have

$$
\begin{aligned}
\tau_1^{-1}\left(\frac{1}{36}\delta_1^4-\frac{1}{8}\kappa_1^2\right)-\tau_2^{-1}\left(\frac{1}{36}\delta_2^4-\frac{1}{8}\kappa_2^2\right)&=1-1=0\\
\frac{1}{2}\tau_1^{-1}\delta_1\kappa_1-\frac{1}{2}\tau_2^{-1}\delta_2\kappa_2&=\frac{1}{2}(\varphi_1-\varphi_2)\\
\tau_1^{-1}\delta_1^i\kappa_1^j-\tau_2^{-1}\delta_2^i\kappa_2^j&=\begin{cases}=\varphi_1\delta_1^{i-1}\kappa_1^{j-1}-\varphi_2\delta_2^{i-1}\kappa_2^{j-1} & \text{if } i\geq 1,\\ =\tau_1^{-1}\kappa_1^j-\tau_2^{-1}\kappa_2^j\leq\sup\left|\frac{\kappa^2}{\tau}\right|(\kappa_1+\kappa_2) & \text{if } i=0\end{cases},
\end{aligned}
$$

and the same applies to $\tau_1^{-2}\delta_1^{2i}\kappa_1^{2j}$. Together with Lemma 8.3, which implies that

$$E\left[\sup_{|\delta|\leq\bar{\delta},|\kappa|\leq\bar{\kappa}}\left(\frac{1}{\sqrt{n}}\frac{\partial^6 L_n(\delta,\kappa,1)}{\partial\lambda\partial\delta^i\partial\kappa^j}\right)^2\right]\to E\left[\sup_{|\delta|,|\kappa|}\left(\mathcal{G}^{[i,j]}(\delta,\kappa)\right)^2\right]<\infty,$$

we can find $c_1\geq c_2>0$ such that

$$E\left[\sup_{\|(\tau_1,\varphi_1)-(\tau_2,\varphi_2)\|\leq c_2,\tau_1,\tau_2\leq 2c_1}\left(\mathcal{G}_n^d(\tau_1,\varphi_1)-\mathcal{G}_n^d(\tau_2,\varphi_2)\right)^2\right]\leq\frac{\varepsilon^3}{2}. \tag{B24}$$

Then Chebychev's inequality implies that

$$\Pr\left[\sup_{\|(\tau_1,\varphi_1)-(\tau_2,\varphi_2)\|\le c_2,|\tau_1|,|\tau_2|\le 2c_1}\left|\mathcal{G}_n^d(\tau_1,\varphi_1)-\mathcal{G}_n^d(\tau_2,\varphi_2)\right|>\varepsilon\right]$$
$$\le\frac{1}{\varepsilon^2}E\left[\sup_{\|(\tau_1,\varphi_1)-(\tau_2,\varphi_2)\|\le c_2,|\tau_1|,|\tau_2|\le 2c_1}\left(\mathcal{G}_n^d(\tau_1,\varphi_1)-\mathcal{G}_n^d(\tau_2,\varphi_2)\right)^2\right]\le\frac{\varepsilon}{2}.$$

Step 2. Given c_1, we need to find c_3 such that $c_1 \ge c_3 > 0$ and (B16) holds. First, we change (τ,φ) into (δ,κ) for simplicity. For $(\tau,\varphi)\in D^1$, it holds that

$$\frac{1}{36}\delta^4\ge\frac{1}{36}\delta^4-\frac{1}{8}\kappa^2=\tau(\delta,\kappa)\ge c_1,\delta\ge 0,$$

which implies $\delta\ge\sqrt{6}c_1^{\frac{1}{4}}$. Moreover, for all $c_B>0$, there exists a $c_3>0$ such that

$$\begin{aligned}&\left\{(\tau_1,\varphi_1,\tau_2,\varphi_2)\in B^1_{\tau\varphi}\times B^1_{\tau\varphi}:\|(\tau_1,\varphi_1)-(\tau_2,\varphi_2)\|\le c_3,\tau_1,\tau_2\ge c_1\right\}\\&\quad\subset\{(\tau_1,\varphi_1,\tau_2,\varphi_2)\in B^1_{\tau\varphi}\times B^1_{\tau\varphi}:\|(\delta_1,\kappa_1)-(\delta_2,\kappa_2)\|\le c_B,\delta_1,\delta_2\ge\sqrt{6}c_1^{\frac{1}{4}}\}\end{aligned}\qquad\text{(B25)}$$

because $\left\{(\tau,\varphi)\in D^1_{\tau\varphi}:\tau\ge c_1\right\}$ is a compact set, and $\tau(\delta,\kappa)$ and $\varphi(\delta,\kappa)$ are continuous on this set. Therefore, it suffices to find c_B such that $\{\mathcal{G}_n(\delta,\kappa):|\delta|\ge\sqrt{6}c_1^{1/4},(\delta,\kappa)\in A^1_{\delta\kappa}\}$ is stochastically equicontinuous on (B25). To do so, we use Theorem 1 of Andrews (1994). Specifically, we use the notation f for $\mathcal{G}_n(\delta,\kappa)=\frac{1}{\sqrt{n}}\sum_i f(y_i,\delta,\kappa)$ and show that f belongs to the type II class of functions defined in Andrews (1994, p.2270). This is the class of Lipschitz functions in (δ,κ), which is such that

$$|f(.,\delta_1,\kappa_1)-f(.,\delta_2,\kappa_2)|\le M(.)\left(|\delta_1-\delta_2|+|\kappa_1-\kappa_2|\right)$$

for all $(\delta_1,\kappa_1),(\delta_2,\kappa_2)\in A^1_{\delta\kappa},|\delta_1|,|\delta_2|\ge\sqrt{6}c_1^{1/4}$.

Note that

$$\begin{aligned}\frac{1}{\tau_1}\frac{\partial l}{\partial\lambda}(\tau_1,\varphi_1)-\frac{1}{\tau_2}\frac{\partial l}{\partial\lambda}(\tau_2,\varphi_2)&=y^2\left[D_1(\tau_1,\delta_1,\kappa_1)-D_1(\tau_2,\delta_2,\kappa_2)\right]\\&\quad+y\left[D_2(\tau_1,\delta_1,\kappa_1)-D_2(\tau_2,\delta_2,\kappa_2)\right]\\&\quad+\left[D_3(\tau_1,\delta_1,\kappa_1)-D_3(\tau_2,\delta_2,\kappa_2)\right]\\&\quad-\frac{1}{\tau_1}\exp\left[-\frac{e^{\frac{\delta_1^2}{3}-\kappa_1}}{2}(\delta_1+y)^2+\frac{1}{2}y^2+\frac{1}{6}\delta_1^2-\frac{1}{2}\kappa_1\right]\\&\quad+\frac{1}{\tau_2}\exp\left[-\frac{e^{\frac{\delta_2^2}{3}-\kappa_2}}{2}(\delta_2+y)^2+\frac{1}{2}y^2+\frac{1}{6}\delta_2^2-\frac{1}{2}\kappa_2\right],\end{aligned}\qquad\text{(B26)}$$

where

$$D_1(\tau,\delta,\kappa)=\frac{1}{2}\tau^{-1}e^{\kappa-\frac{\delta^2}{3}}+\frac{1}{2}\frac{\delta^2}{\tau},\quad D_2(\tau,\delta,\kappa)=-\frac{\delta}{\tau}\ \text{ and }\ D_3(\tau,\delta,\kappa)=-\frac{1}{2}\tau^{-1}\left(e^{\kappa-\frac{\delta^2}{3}}-\delta^2\right)$$

so that D_1, D_2 and D_3 are all Lipschitz in (δ,κ) for $(\delta,\kappa)\in A^1_{\delta\kappa}$ and $\tau=\tau(\delta,\kappa)$. And for the last term in (B26), the mean value theorem implies that

$$\begin{aligned}
&-\frac{1}{\tau_1}\exp\left[-\frac{e^{\frac{\delta_1^2}{3}-\kappa_1}}{2}(\delta_1+y)^2+\frac{1}{2}y^2+\frac{1}{6}\delta_1^2-\frac{1}{2}\kappa_1\right]\\
&+\frac{1}{\tau_2}\exp\left[-\frac{e^{\frac{\delta_2^2}{3}-\kappa_2}}{2}(\delta_2+y)^2+\frac{1}{2}y^2+\frac{1}{6}\delta_2^2-\frac{1}{2}\kappa_2\right]\\
=&\exp\left[-\frac{e^{\frac{\tilde\delta^2}{3}-\tilde\kappa}}{2}(\tilde\delta+y)^2+\frac{1}{2}y^2+\frac{1}{6}\tilde\delta^2-\frac{1}{2}\tilde\kappa\right]\left\{\frac{1}{\tilde\tau^2}(\tau_1-\tau_2)\right.\\
&+\frac{1}{3\tilde\tau}\left[e^{\frac{\tilde\delta^2}{3}-\tilde\kappa}\left(\tilde\delta^3+3\tilde\delta+\tilde\delta y^2+2\tilde\delta^2y+3y\right)-\tilde\delta\right](\delta_1-\delta_2)\\
&\left.+\frac{1}{2\tilde\tau}\left[1-e^{\frac{\tilde\delta^2}{3}-\tilde\kappa}(\tilde\delta+y)^2\right](\kappa_1-\kappa_2)\right\}. \qquad \text{(B27)}
\end{aligned}$$

In addition,

$$\begin{aligned}
|\tau_1-\tau_2| &= \left|\frac{1}{36}\delta_1^4-\frac{1}{8}\kappa_1^2-\frac{1}{36}\delta_2^4+\frac{1}{8}\kappa_2^2\right|\\
&=\left|\frac{1}{36}\left(\delta_1^2+\delta_2^2\right)(\delta_1+\delta_2)(\delta_1-\delta_2)-\frac{1}{8}\left(\kappa_1+\kappa_2\right)\left(\kappa_1-\kappa_2\right)\right|\\
&\le \frac{1}{9}\bar\delta^3\left|\delta_1-\delta_2\right|+\frac{\bar\kappa}{4}\left|\kappa_1-\kappa_2\right|. \qquad \text{(B28)}
\end{aligned}$$

Moreover

$$\exp\left[-\frac{e^{\frac{\delta^2}{3}-\kappa}}{2}(\delta+y)^2+\frac{1}{2}y^2+\frac{1}{6}\delta^2-\frac{1}{2}\kappa\right]\le g^*(y), \qquad \text{(B29)}$$

where

$$g^*(y)=\exp\left[-\frac{e^{-\bar\kappa}}{2}(2\bar\delta|y|+y^2)+\frac{1}{2}y^2+\frac{1}{6}\bar\delta^2+\frac{1}{2}\bar\kappa\right].$$

Combining (B26), (B27), (B28) and (B29), we will have

$$\frac{1}{\tau_1}\frac{\partial l}{\partial\lambda}(\tau_1,\varphi_1)-\frac{1}{\tau_2}\frac{\partial l}{\partial\lambda}(\tau_2,\varphi_2)\le\left(g^*(y)+1\right)\left\{a_1+a_2|y|+a_3y^2\right\}\left(|\delta_1-\delta_2|+|\kappa_1-\kappa_2|\right).$$

But since

$$E\left[\left(g^*(y)+1\right)\left\{a_1+a_2|y|+a_3y^2\right\}\right]<\infty,$$

f will be Lipschitz with $M(y)=(g^*(y)+1)\left(a_1+a_2|y|+a_3y^2\right)$ for some constants a_1, a_2 and a_3. To apply Theorem 1 of Andrews (1994), we need to check Assumptions A, B, and C. Assumption A: the class of functions f satisfies Pollard's entropy condition with some envelope $\bar M$. This is satisfied with $\bar M=1\vee\sup|f|\vee M(.)$ by Theorem 2 of Andrews (1994) because f is Lipschitz.

In turn, Assumption B:

$$\lim_{n\to\infty}\sup \frac{1}{n}\sum_{i=1}^{n} E\bar{M}^{2+\upsilon}(y_i) < \infty \text{ for some } \upsilon > 0,$$

is also satisfied because y_i is a standard normal random variable. Finally, Assumption C: $\{y_i\}$ is an m-dependent triangular array of r.v's holds because $\{y_i\}$ is *iid*. Stochastic equicontinuity of f follows from Theorem 1 of Andrews (1994). Thus, for given $\varepsilon > 0$, we can find c_B such that (B16) holds.

In sum, the results hold by virtue of (B17) and (B18). □

Step 3

Lemma 3 $\sup_{d\in D^1} LR_n^d(d) = \sup_{d\in D^1} LM_n^d(d) + o_p(1) = \sup_{(\tau,\varphi)\in D^1_{\tau\varphi}} \frac{\left[\mathcal{G}_n^d(\tau,\varphi)\right]_-^2}{V^d(\tau,\varphi)} + o_p(1)$.

Proof. Since

$$\left|\sup_{d\in D^1} LR_n^d(d) - \sup_{d\in D^1} LM_n^d(d)\right| \leq \sup_{(\tau,\varphi)\in D^1_{\tau\varphi}} \left|\sup_{\eta:(\eta,\tau,\varphi)\in D^1} LR_n^d(\eta,\tau,\varphi) - \sup_{\eta:(\eta,\tau,\varphi)\in D^1} LM_n^d(\eta,\tau,\varphi)\right|,$$

it suffices to show that

$$\sup_{\eta:(\eta,\tau,\varphi)\in D^1} LR_n^d(\eta,\tau,\varphi) = \sup_{\eta:(\eta,\tau,\varphi)\in D^1} LM_n^d(\eta,\tau,\varphi) + o_p(1). \tag{B30}$$

Expression (B30) follows from Andrews (2001). To see this, we need to check his assumptions. Let

$$l^d(\eta,\tau,\varphi) = l(\delta(\tau,\varphi),\kappa(\tau,\varphi),\lambda(\eta,\tau,\varphi))$$

denote the log-likelihood of y_i written in $d \in D^1$. The null hypothesis is $H_0 : \eta = 0$ and (τ,φ) is the nuisance parameter that only appears under the alternative. Let

$$LR_n^d(\hat{\eta}_{\tau\varphi},\tau,\varphi) = \sup_{\eta:(\eta,\tau,\varphi)\in D^1} LR_n^d(\eta,\tau,\varphi).$$

To verify Assumption 1, namely $\hat{\eta}_{\tau\varphi} = o_{p,\tau\varphi}(1)$, let $l_0^d(d) = E\left[l^d(1,\tau,\varphi)\right]$. Invoking Lemma 6, we have

$$\sup_{d\in D^1}\left|\frac{1}{n}L_n^d(d) - l_0^d(0,\tau,\varphi)\right| \leq \sup_{\theta\in\Theta}\left|\frac{1}{n}L_n(\theta) - l_0(\theta)\right| \xrightarrow{p} 0 \tag{B31}$$

(i.e. uniform convergence). Moreover, for all $\epsilon > 0$,

$$l_0^d(d) > \sup_{\eta>\epsilon, d\in \mathrm{cl}(D^1)} l_0^d(d) \tag{B32}$$

(i.e. well separated maximum), which follows from the fact that $\eta = 1$ is the unique maximizer (note that $(1-\lambda) \leq \max\{|\delta|,|\kappa|\}$), $l_0^d(d)$ is continuous and $\mathrm{cl}(D^1)$ is compact. As a result, Lemma A1 in Andrews (1993) implies that we have $\hat{\eta}_{\tau\varphi} = o_{p,\tau\varphi}(1)$.

Assumption 2* holds with $B_T = \sqrt{n}$ using Andrews (2001) notation, see Lemma 1. Assumption 3* holds by Lemma 2. Assumption 4 is implied by Assumptions 1, 2* and 3. Assumption 5 is satisfied for $B_T = b_T = \sqrt{n}$ and $\Lambda = \mathbb{R}^-$. Assumption 6 holds because $\mathbb{R}^-$ is convex. Assumptions 7 and 8 hold with $\Lambda_\beta = \mathbb{R}^-$ and with the fact that δ and ψ are absent in our setting. Assumptions 9 and 10 are satisfied. Assumptions 1o and 4o hold trivially because the restricted estimator is $\eta = 0$ and therefore not random.

By Theorem 4 and the remark at the bottom of p. 719 of Andrews (2001), it follows that (B30) holds.

Step 4

In this step, we show that

$$\sup_{\vartheta\in\Theta'} 2\left[\mathcal{L}_n(\vartheta) - \mathcal{L}_n(0,0,1)\right] = \frac{1}{n}\sup_{\vartheta\in\Theta'\setminus(0,0,1)} \frac{\left(\min\left\{\partial\mathcal{L}_n(\delta,\varkappa,1)/\partial\lambda, 0\right\}\right)^2}{V(\delta,\varkappa)} + o_p(1),$$

where we use the notation $\mathcal{L}_n$ for the log-likelihood indexed by ϑ, whereas L_n is the log-likelihood indexed by θ. First, by the results in Step 3, we have

$$\sup_{d\in D^k} LR_n^d(d) = \sup_{(\tau,\varphi)\in D^k_{\tau\varphi}} \frac{\left\lfloor \mathcal{G}_n^d(\tau,\varphi)\right\rfloor_-^2}{V^d(\tau,\varphi)} + o_p(1).$$

Noticing also that

$$\sup_{d\in D^k} LR_n^d(d) = \sup_{\theta\in A^k} LR_n(\theta) \quad \text{and} \quad \sup_{(\tau,\varphi)\in D^k_{\tau m}} \frac{\left\lfloor \mathcal{G}_n^d(\tau,\varphi)\right\rfloor_-^2}{V^d(\tau,\varphi)} = \sup_{(\delta,\kappa)\in A^k_{\tau m}} \frac{\left\lfloor \mathcal{G}_n(\delta,\kappa)\right\rfloor_-^2}{V(\delta,\kappa)},$$

we will have that

$$\begin{aligned}
\sup_{\theta\in\mathcal{P}_b} LR_n(\theta) &= \max_{k\le 16}\sup_{d\in D^k} LR_n^d(d) = \max_{k\le 16}\sup_{(\delta,\kappa)\in A^k_{\tau m}} \frac{\left\lfloor \mathcal{G}_n(\delta,\kappa)\right\rfloor_-^2}{V(\delta,\kappa)} + o_p(1) \\
&= \sup_{(\delta,\kappa):(\delta,\kappa,1)\in\mathcal{P}_b} \frac{\left\lfloor \mathcal{G}_n(\delta,\kappa)\right\rfloor_-^2}{V(\delta,\kappa)} + o_p(1).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\sup_{\vartheta\in\mathcal{P}_b'} 2\left(\mathcal{L}_n(\vartheta) - \mathcal{L}_n(0,0,1)\right) &= \sup_{\theta\in\mathcal{P}_b} LR_n(\theta) = \sup_{(\delta,\kappa):(\delta,\kappa,1)\in\mathcal{P}_b} \frac{\left\lfloor \mathcal{G}_n(\delta,\kappa)\right\rfloor_-^2}{V(\delta,\kappa)} + o_p(1) \\
&= \sup_{(\delta,\varkappa):(\delta,\varkappa,1)\in\mathcal{P}_b'} \frac{\left\lfloor \mathcal{G}_n(\delta,\varkappa)\right\rfloor_-^2}{V(\delta,\varkappa)}.
\end{aligned}$$

Finally, the asymptotic distributions of the LM tests follow from the continuous mapping theorem. □

C Detailed proof of Proposition 6

Constant μ and σ^2

We first consider the simple case in which we estimate both the unconditional mean and variance parameters, say μ and σ^2, respectively, under the additional assumption that they are constants. Specifically, letting $y = \sqrt{\sigma^2}z + \mu$ and $z \sim \text{MixN}(0,1)$, we have that the pdf of y is simply given by

$$f_Y(y) = \frac{1}{\sqrt{\sigma^2}} f_Z\left(\frac{y-\mu}{\sqrt{\sigma^2}}\right),$$

so that the contribution of observation y to the log-likelihood, $\ell(\mu, \sigma^2, \delta, \varkappa, \lambda; y)$, will be given by

$$k - \frac{1}{2}\log\sigma^2 + \log\left\{\frac{\lambda}{\sqrt{\sigma_1^{*2}}}\exp\left[-\frac{1}{2\sigma_1^{*2}}\left(\frac{y-\mu}{\sqrt{\sigma^2}} - \mu_1^*\right)^2\right] + \frac{1-\lambda}{\sqrt{\sigma_2^{*2}}}\exp\left[-\frac{1}{2\sigma_2^{*2}}\left(\frac{y-\mu}{\sqrt{\sigma^2}} - \mu_2^*\right)^2\right]\right\},$$

where k is an integration constant and

$$\mu_1^* = \frac{\delta(1-\lambda)}{\sqrt{1+\lambda(1-\lambda)\delta^2}}, \quad \mu_2^* = -\frac{\lambda}{1-\lambda}\mu_1^*,$$

$$\sigma_1^{*2} = \frac{1}{\left[1+\lambda(1-\lambda)\delta^2\right]\left[\lambda + (1-\lambda)\exp(\varkappa)\right]} \quad \text{and} \quad \sigma_2^{*2} = \exp(\varkappa)\sigma_1^{*2}.$$

Subtest in $\mathcal{P}_a$ We consider the reparametrization in (3) and define

$$L_n(\mu, \sigma^2, \delta, \kappa, \lambda) = \frac{1}{n}\sum_{i=1}^{n} l_i(\mu, \sigma^2, \delta, \kappa, \lambda),$$

with $l_i(\mu, \sigma^2, \delta, \kappa, \lambda) = \ell(\mu, \sigma^2, \delta, \kappa - (2\lambda - 1)\delta^2/3, \lambda; y_i)$.

To shorten notation, let $\rho = (\phi, \theta)$ with $\phi = (\mu, \sigma^2)$ and $\theta = (\delta, \kappa, \lambda)$. Let $\phi_0 = (\mu_0, \sigma_0^2)$ denote the true value of the parameter ϕ. Next, define

$$LR_n(\mu, \sigma^2, \delta, \kappa, \lambda) = 2\left[L_n(\mu, \sigma^2, \delta, \kappa, \lambda) - L_n(\mu_0, \sigma_0^2, 0, 0, \lambda)\right] \tag{C1}$$

and

$$\rho_{n,r}^{LR} = \underset{\rho\in\Phi\times\{0\}^2\times[1/2,1]}{\arg\max} LR(\rho), \quad \rho_{n,u}^{LR} = \underset{\rho\in\Phi\times\mathcal{P}}{\arg\max} LR(\rho),$$

where $\mathcal{P}$ can be replaced by $\mathcal{P}_{a,1}, \mathcal{P}_{a,2}, \mathcal{P}_{a,3}$ as needed, and Φ denotes the feasible parameter space of (μ, σ^2). Then, it is easy to verify that $\rho_{n,r}^{LR} = (\phi_{n,r}, 0, 0, \lambda_{n,r})$ with

$$\phi_{n,r} = (\mu_{n,r}, \sigma_{n,r}^2) = \left[\frac{1}{n}\sum_{i=1}^{n} y_i, \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu_{n,r})^2\right],$$

which provide the restricted maximum likelihood estimators of ϕ.

Letting

$$LM_n^{a,\phi}(\phi) = 2\left(\frac{1}{\sigma_0}\frac{H_{1,n}}{\sqrt{n}}\right)\sqrt{n}(\mu-\mu_0) + 2\left(\frac{1}{2\sigma_0^2}\frac{H_{2,n}}{\sqrt{n}}\right)\sqrt{n}(\sigma^2-\sigma_0^2) - \frac{1}{\sigma_0^2}n(\mu-\mu_0)^2 - \frac{1}{2\sigma_0^4}n\left(\sigma^2-\sigma_0^2\right)^2, \tag{C2}$$

where

$$H_{1,n} = \sum_{i=1}^n h_{1i} = \sum_{i=1}^n \frac{y_i-\mu_0}{\sqrt{\sigma_0^2}} \quad \text{and} \quad H_{2,n} = \sum_{i=1}^n h_{2i} = \sum_{i=1}^n \frac{(y_i-\mu_0)^2-\sigma_0^2}{\sigma_0^2}.$$

Moreover, in the sequel $LM_n^a(\theta;\phi_0)$ will coincide with (15) if we replace y_i with $(y_i-\mu_0)/\sqrt{\sigma_0^2}$. As in the proof of Proposition 1, we have the following five steps:

1. For all sequences of $\rho_n = (\phi_n, \delta_n, \kappa_n, \lambda_n)$ with $(\phi_n, \delta_n, \kappa_n) \xrightarrow{p} (\phi_0, 0, 0)$, we have that

$$LR_n(\rho_n) = LM_n^a(\theta_n) + LM_n^{a,\phi}(\phi_n) + o_p[h_n^\theta(\theta_n)] + o_p[h_n^\phi(\phi_n)],$$

where $h_n^\phi(\phi) = \max\left\{1, n(\mu-\mu_0)^2, n(\sigma^2-\sigma_0^2)^2\right\}$ and

$$h_n^\theta(\theta) = \max\left\{1, n(1-\lambda)^2\delta^8, n(1-\lambda)^2\delta^2\kappa^2, n(1-\lambda)^2\kappa^4\right\}.$$

2. For $\phi_n = (\mu_n^{LM}, \sigma_n^{2LM}) \in \arg\max_{\phi\in\Phi} LM_n^{a,\phi}(\phi)$, we have that $\phi_n^{LM} = \phi_0 + o_p(1)$ and $h_n^\phi(\phi_n^{LM}) = O_p(1)$; and also define $\theta_n^{LM} = (\delta_n^{LM}, \kappa_n^{LM}, \lambda_n^{LM}) \in \arg\max_{\theta\in\Theta} LM_n^a(\theta)$, we have that $(\delta_n^{LM}, \kappa_n^{LM}) = o_p(1)$ and $h_n^\theta(\theta_n^{LM}) = O_p(1)$.

3. For $\rho_{n,u}^{LR} = (\phi_{n,u}^{LR}, \delta_{n,u}^{LR}, \kappa_{n,u}^{LR}, \lambda_{n,u}^{LR}) \in \arg\max_{\phi\in\Phi\times\mathcal{P}} LR_n(\rho)$, we have that

$$(\phi_{n,u}^{LR} - \phi_0, \delta_{n,u}^{LR}, \kappa_{n,u}^{LR}) \xrightarrow{p} 0$$

and $h(\rho_{n,u}^{LR}) = O_p(1)$.

4. Then, we prove that $LR_n(\rho_{n,r}^{LR}) - LR_n(\rho_{n,u}^{LR}) = LM_n^a(\theta_n^{LM}) + o_p(1)$.

5. Finally, show that the test is the same as before, but with y_i replaced by $(y_i - \mu_{n,r})/\sigma_{n,r}$.

Before going into the details of these steps, let us emphasize that the main difference is in Step 1, which shows that in the Taylor expansion the cross terms (T_3 defined below) of ϕ and θ are negligible, and thus we can consider the two parts separately. Step 2-4 are almost the same as before.

Step 1: Consider a sequence $\rho_n = (\phi_n, \delta_n, \kappa_n, \lambda_n)$ with $(\phi_n, \delta_n, \kappa_n) \xrightarrow{p} (\phi_0, 0, 0)$. Let

$$L_n^{[k_1,k_2,k_3,k_4]} = \frac{1}{k_1!k_2!k_3!k_4!}\frac{\partial^{k_1+k_2+k_3+k_4}L_n(\rho)}{\partial\mu^{k_1}\left(\partial\sigma^2\right)^{k_2}\partial\delta^{k_3}\partial\kappa^{k_4}}\bigg|_{\rho_{n,0}}$$

where $\rho_{n,0} = (\phi_0, 0, 0, \lambda_n)$ and

$$\triangle_n^{[k_1,k_2,k_3,k_4]} = \frac{1}{k_1!k_2!k_3!k_4!}\left[\left.\frac{\partial^{k_1+k_2+k_3+k_4}L_n(\rho)}{\partial\mu^{k_1}(\partial\sigma^2)^{k_2}\partial\delta^{k_3}\partial\kappa^{k_4}}\right|_{(\tilde{\phi}_n,\tilde{\delta}_n,\tilde{\kappa}_n,\lambda_n)} - \left.\frac{\partial^{k_1+k_2}L_n(\rho)}{\partial\delta^{k_1}\partial\kappa^{k_2}}\right|_{\rho_{n,0}}\right],$$

with $(\tilde{\phi}_n, \tilde{\delta}_n, \tilde{\kappa}_n)$ between $(\phi_0, 0, 0)$ and $(\phi_n, \delta_n, \kappa_n)$. Consider the following eighth-order Taylor expansion,

$$\begin{aligned}\frac{1}{2}LR_n(\rho_n) &= L_n(\mu_n, \sigma_n^2, \delta_n, \kappa_n, \lambda_n) - L_n(\mu_0, \sigma_0^2, 0, 0, \lambda_n)\\ &= T_{1n}(\theta_n; \phi_0) + T_{2n}(\phi_n; \phi_0) + T_{3n}(\rho_n; \mu_0, \sigma_0^2) + \Delta_n,\end{aligned}$$

where

$$\begin{aligned}T_{1n}(\theta_n;\phi_0) &= \sum_{k_3+k_4\le 8} L_n^{[0,0,k_3,k_4]}\delta_n^{k_3}\kappa_n^{k_4},\\ T_{2n}(\phi_n;\phi_0) &= \sum_{k_1+k_2\le 8} L_n^{[k_1,k_2,0,0]}(\mu_n-\mu_0)^{k_1}\left(\sigma_n^2-\sigma_0^2\right)^{k_2},\\ T_{3n}(\rho_n;\phi_0) &= \sum_{\substack{k_1+k_2+k_3+k_4\le 8\\ k_1+k_2\ge 1,\ k_3+k_4\ge 1}} L_n^{[k_1,k_2,k_3,k_4]}(\mu_n-\mu_0)^{k_1}\left(\sigma_n^2-\sigma_0^2\right)^{k_2}\delta_n^{k_3}\kappa_n^{k_4} \quad\text{and}\\ \Delta_n &= \sum_{k_1+k_2+k_3+k_4=8} \Delta_n^{[k_1,k_2,k_3,k_4]}(\mu_n-\mu_0)^{k_1}\left(\sigma_n^2-\sigma_0^2\right)^{k_2}\delta_n^{k_3}\kappa_n^{k_4}\end{aligned}$$

First, we will show that $T_{3n}(\rho_n;\phi_0) = o_p[h_n^{\theta}(\theta_n)] + o_p[h_n^{\phi}(\phi_n)]$. Specifically, for $(k_1,k_2) \in \{(1,0),(0,1)\}$ and $(k_3,k_4) \in \{(k,0) : k \le 4\} \cup \{(0,k) : k \le 2\} \cup \{(1,1)\}$, we can easily check that

$$E[l^{[k_1,k_2,k_3,k_4]}(\rho_0)] = 0 \text{ and } E\{[l^{[k_1,k_2,k_3,k_4]}(\rho_0)]^2\} < \infty,$$

which means that

$$\frac{\sqrt{n}}{n}\left.\frac{\partial^{k_1+k_2+k_3+k_4}L_n(\rho)}{\partial\mu^{k_1}(\partial\sigma^2)^{k_2}\partial\delta^{k_3}\partial\kappa^{k_4}}\right|_{\rho_0} = O_p(1). \tag{C3}$$

Therefore, we will have that the (k_1,k_2,k_3,k_4) term is such that

$$\begin{aligned}L_n^{[k_1,k_2,k_3,k_4]}(\mu_n-\mu_0)^{k_1}\left(\sigma_n^2-\sigma_0^2\right)^{k_2}\delta_n^{k_3}\kappa_n^{k_4} &= \left(\frac{\sqrt{n}}{n}\left.\frac{\partial^{k_1+k_2+k_3+k_4}L_n(\rho)}{\partial\mu^{k_1}(\partial\sigma^2)^{k_2}\partial\delta^{k_3}\partial\kappa^{k_4}}\right|_{\rho_0}\right)\\ &\quad\times\left[\sqrt{n}(\mu_n-\mu_0)^{k_1}\left(\sigma_n^2-\sigma_0^2\right)^{k_2}\right]\delta_n^{k_3}\kappa_n^{k_4}\\ &= o_p[h_n^{\phi}(\phi_n)],\end{aligned}$$

where the last equality follows from (C3) and the fact that $\delta_n^{k_3}\kappa_n^{k_4} = o_p(1)$. As for the remaining terms in T_{3n}, we have either: a) $k_1 + k_2 \ge 2$ so that

$$n(\mu_n-\mu_0)^{k_1}\left(\sigma_n^2-\sigma_0^2\right)^{k_2}\delta_n^{k_3}\kappa_n^{k_4} = o_p[h_n^{\phi}(\phi_n)], \tag{C4}$$

or b) $(k_3,k_4)\in\{(k,0):k>4\}\cup\{(0,k):k>2\}\cup\{(k,k'):k,k'>1\}$, so that

$$\begin{aligned} L_n^{[k_1,k_2,k_3,k_4]}\left(\mu_n-\mu_0\right)^{k_1}\left(\sigma_n^2-\sigma_0^2\right)^{k_2}\delta_n^{k_3}\kappa_n^{k_4} &= \left[\frac{1}{n}\sum_{i=1}^{n}g(y_i)\right]n\left(\mu_n-\mu_0\right)^{k_1}\left(\sigma_n^2-\sigma_0^2\right)^{k_2} \\ &\quad \times(1-\lambda_n)\delta_n^{k_3}\kappa_n^{k_4} \\ &= o_p[h_n^\theta(\theta_n)], \end{aligned}$$

where $g(y)=l_n^{[k_1,k_2,k_3,k_4]}(\rho_{n0})/(1-\lambda_n)$ is square integrable. In this case, the last equality follows from

$$\sqrt{n}\left(\mu_n-\mu_0\right)^{k_1}\left(\sigma_n^2-\sigma_0^2\right)^{k_2}\sqrt{n}(1-\lambda_n)\delta_n^{k_3}\kappa_n^{k_4}=o_p[h_n^\theta(\theta_n)]. \tag{C5}$$

Secondly, we have to show that $T_{2n}=LM_n^{a,\phi}(\phi_n;\phi_0)+o_p[h_n^\phi(\phi_n)]$. Invoking Rotnitzky et al (2000), we will have that

$$\frac{1}{n}L_n^{[2,0,0,0]}=-\frac{1}{2\sigma_0^2}+O_p(n^{-\frac{1}{2}}),\ \ \frac{1}{n}L_n^{[0,2,0,0]}=-\frac{1}{4\sigma_0^2}+O_p(n^{-\frac{1}{2}})\ \text{ and }\ \frac{1}{n}L_n^{[1,1,0,0]}=O_p(n^{-\frac{1}{2}}).$$

Therefore

$$\begin{aligned} \sum_{k_1+k_2=2}L_n^{[k_1,k_2,0,0]}\left(\mu_n-\mu_0\right)^{k_1}\left(\sigma_n^2-\sigma_0^2\right)^{k_2} &= \sum_{k_1+k_2=2}\frac{1}{n}L_n^{[k_1,k_2,0,0]}n\left(\mu_n-\mu_0\right)^{k_1}\left(\sigma_n^2-\sigma_0^2\right)^{k_2} \\ &= -\frac{1}{2\sigma_0^2}n\left(\mu_n-\mu_0\right)^2-\frac{1}{4\sigma_0^2}n\left(\sigma_n^2-\sigma_0^2\right)^2+o_p[h_n^\phi(\phi_n)]. \end{aligned}$$

For $k_1+k_2>2$, we have $\frac{1}{n}L_n^{[k_1,k_2,0,0]}=O_p(1)$ and $n\left(\mu_n-\mu_0\right)^{k_1}\left(\sigma_n^2-\sigma_0^2\right)^{k_2}=o_p\left[h_n^\phi(\phi_n)\right]$.

Third, we have to show that $T_{1n}=LM_n^a(\theta_n)+o_p[h_n^\theta(\theta_n)]$. But since this is the same as we did in the proof of Proposition 1, we can omit it.

The last part requires to prove that $\Delta_n^{[k_1,k_2,k_3,k_4]}\left(\mu_n-\mu_0\right)^{k_1}\left(\sigma_n^2-\sigma_0^2\right)^{k_2}\delta_n^{k_3}\kappa_n^{k_4}=o_p(1)$ for $k_1+k_2+k_3+k_4=8$, which is entirely analogous to the proof of Proposition 1.

Step 2: This step is trivial since $\max_{\phi\in\Phi}LM^{a,\phi}(\phi)$ has a closed-form solution with probability approaching one.

Step 3: Following the proof of Proposition 1, we can first show that $\left(\phi_{n,u}^{LR},\delta_{n,u}^{LR},\kappa_{n,u}^{LR}\right)\xrightarrow{p}(\phi_0,0,0)$. Next, we can also show that $h_n^\theta(\theta_{n,u}^{LR})=O_p(1)$ and $h_n^\phi(\phi_{n,u}^{LR})=O_p(1)$ by an argument analogous to Lemma 3 in Amengual, Bei and Sentana (2023).

Step 4: It follows from the same argument as in the corresponding proof of Proposition 1.

Step 5: Simplify $LM_n^a(\theta_n^{LM})$ as in the proof of Proposition 1. Then by the stochastic equicontinuity of the test statistic in ϕ, we can replace ϕ by $\phi_{n,r}$.

Subtest in $\mathcal{P}_b$ Here we use the reparametrization of Proposition 3 involving (η,τ,φ). In terms of Andrews (2001) notation, we have

$$\beta_1=\eta,\ \ \pi=(\tau,\varphi)\ \text{ and }\ \psi=(\mu,\sigma^2).$$

We show that we do not need to adjust for parameter uncertainty by verifying Assumption 7 of Andrews (2001), which guarantees that there is no cross term of ϕ and η in the quadratic approximation. Let

$$\begin{aligned}
LR_n^d(\mu,\sigma^2,\eta,\tau,\varphi) =& LR_n[\mu,\sigma^2,\delta(\tau,\varphi),\kappa(\tau,\varphi),\lambda(\eta,\tau,\varphi)],\\
LM_n^d(\mu,\sigma^2,\eta,\tau,\varphi) =& 2\mathcal{G}_n(\tau,\varphi)\sqrt{n}\eta - V(\tau,\varphi)n\eta^2 + LM_n^{\phi}(\phi),\\
R_n^d(\mu,\sigma^2,\eta,\tau,\varphi) =& LR_n^d(\mu,\sigma^2,\eta,\tau,\varphi) - LM_n^d(\mu,\sigma^2,\eta,\tau,\varphi),
\end{aligned}$$

where $LR_n(\mu,\sigma^2,\delta,\kappa,\lambda)$ is defined in (C1) and $LM_n^{\phi}(\phi)$ in (C2). We need to show that for all sequences $(\mu_n,\sigma_n^2,\eta_n,\tau_n,\varphi_n)$ with $(\mu_n-\mu_0,\sigma_n^2-\sigma_0^2,\eta_n)\xrightarrow{p}0$, it holds that

$$R_n(\mu_n,\sigma_n^2,\eta_n,\tau_n,\varphi_n) = o_p\left\{\max[n\eta_n^2, n(\mu_n-\mu_0), n(\sigma_n^2-\sigma_0^2)^2]\right\}. \tag{C6}$$

To see this, we can modify the proof of Proposition 3. Let $\rho_n = (\mu_n,\sigma_n^2,\delta_n,\kappa_n,\lambda_n)$ with $\delta_n=\delta(\tau_n,\varphi_n)$, $\kappa_n=\kappa(\tau_n,\varphi_n)$ and $\lambda_n=\lambda(\eta_n,\tau_n,\varphi_n)$. A third-order Taylor expansion gives

$$\begin{aligned}
L(\mu_n,\sigma_n^2,\delta_n,\kappa_n,\lambda_n) - L(\mu_0,\sigma_0^2,\delta_n,\kappa_n,1) = T_{1n}(\rho_n;\phi_0) + T_{2n}(\rho_n;\phi_0)\\
+ T_{3n}(\rho_n;\phi_0) + T_{4n}(\rho_n;\phi_0),
\end{aligned}$$

where

$$T_{1n}(\rho_n;\phi_0) = \frac{\partial L(\rho_{n0})}{\partial\lambda}(\lambda_n-1) + \frac{1}{2}\frac{\partial^2 L(\rho_{n0})}{\partial\lambda^2}(\lambda_n-1)^2 + \frac{1}{3!}\frac{\partial^3 L(\tilde{\rho}_n)}{\partial\lambda^3}(\lambda_n-1)^3.$$

$$T_{2n}(\rho_n;\phi_0) = \sum_{i+j\le 2}\frac{1}{i!j!}\frac{\partial^{i+j}L(\rho_{n0})}{\partial\mu^i\partial(\sigma^2)^j}(\mu_n-\mu_0)^i(\sigma_n^2-\sigma_0^2)^j + \sum_{i+j=3}\frac{1}{i!j!}\frac{\partial^3 L(\tilde{\rho}_n)}{\partial\mu^i\partial(\sigma^2)^j}(\mu_n-\mu_0)^i(\sigma_n^2-\sigma_0^2)^j$$

and

$$\begin{aligned}
T_{3n}(\rho_n;\phi_0) =& \frac{\partial^2 L(\rho_{n0})}{\partial\lambda\partial\mu}(\lambda_n-1)(\mu_n-\mu_0) + \frac{\partial^2 L(\rho_{n0})}{\partial\lambda\partial\sigma^2}(\lambda_n-1)(\sigma_n^2-\sigma_0^2)\\
&+ \frac{1}{2}\frac{\partial^3 L(\tilde{\rho}_n)}{\partial\lambda^2\partial\mu}(\lambda_n-1)^2(\mu_n-\mu_0) + \frac{1}{2!2!}\frac{\partial^3 L(\tilde{\rho}_n)}{\partial\lambda^2\partial\sigma^2}(\lambda_n-1)^2(\sigma_n^2-\sigma_0^2),
\end{aligned}$$

$$T_{4n} = \sum_{j+k=2}\frac{1}{j!k!}\left\{\frac{1}{n}\frac{\partial^3 L(\tilde{\rho}_n)}{\partial\lambda\partial\mu^j\partial(\sigma^2)^k}\right\}n(\mu_n-\mu_0)^j(\sigma_n^2-\sigma_0^2)^k(\lambda_n-1)$$

with $\tilde{\rho}_n = (\tilde{\mu}_n,\tilde{\sigma}_n^2,\delta_n,\kappa_n,\tilde{\lambda}_n)$ between $(\mu_n,\sigma_n^2,\delta_n,\kappa_n,\lambda_n)$ and $\rho_{n0}=(\mu_0,\sigma_0^2,\delta_n,\kappa_n,1)$. We can show that

$$2T_{1n}(\rho_n;\phi_0) = 2\mathcal{G}_n(\tau_n,m_n)\sqrt{n}\eta_n - V(\tau_n,m_n)n\eta_n^2 + o_p(n\eta_n^2) \tag{C7}$$

using the same argument as in Proposition 3. Moreover, it is straightforward to show that

$$2T_{2n}(\phi_n;\phi_0) = LM_n^{\phi}(\phi_n) + o_p\left[n\left(\sigma_n^2-\sigma_0^2\right)^2 + n\left(\mu_n-\mu_0\right)^2\right] \tag{C8}$$

We can also show that

$$\begin{aligned} T_{3n}(\rho_n;\phi_0) &= \left\{\frac{1}{\sqrt{n}}\frac{\partial^2 L(\rho_{n0})}{\partial\lambda\partial\mu}\right\}\left[\sqrt{n}(\mu_n-\mu_0)\right](\lambda_n-1) \\ &+\left\{\frac{1}{\sqrt{n}}\frac{\partial^2 L(\rho_{n0})}{\partial\lambda\partial\sigma^2}\right\}\left[\sqrt{n}(\sigma_n^2-\sigma_0^2)\right](\lambda_n-1) \\ &-\frac{1}{2}\left\{\frac{1}{n}\tau_n^{-1}\frac{\partial^3 L(\tilde{\rho}_n)}{\partial\lambda^2\partial\mu}\right\}\left[n(\mu_n-\mu_0)\eta_n\right](\lambda_n-1) \\ &-\frac{1}{4}\left\{\frac{1}{n}\tau_n^{-1}\frac{\partial^3 L(\tilde{\rho}_n)}{\partial\lambda^2\partial\sigma^2}\right\}\left[n(\sigma_n^2-\sigma_0^2)\eta_n\right](\lambda_n-1) \\ &= o_p[n(\mu_n-\mu_0)^2+n\left(\sigma_n^2-\sigma_0^2\right)^2+n\eta_n^2], \end{aligned} \tag{C9}$$

where the first equality follows from $\eta_n=(1-\lambda_n)\tau_n$ and the second one follows from Lemma 8 and $\lambda_n \xrightarrow{p} 1$. The result relative to $T_{4,n}$ is easy, as $\lambda_n \to 1$ and $n(\mu_n-\mu_0)^j(\sigma_n^2-\sigma_0^2)^k = O[n(\mu_n-\mu_0)^2+n\left(\sigma_n^2-\sigma_0^2\right)^2]$, so that

$$T_{4n}=o_p[n(\mu_n-\mu_0)^2+n\left(\sigma_n^2-\sigma_0^2\right)^2]. \tag{C10}$$

Combining the results in (C7), (C8), (C9) and (C10), we finally prove (C6).

General μ and σ^2

Let us now consider the general case in which the conditional mean and variance are parametric functions of another observable vector X.

In this context, let $W_t=(Y_t,X_t)$ and assume that

$$f_{Y_t|(X_t,W^{t-1})}(y|x,w^{t-1})=f_{Y_t|X_t}(y|x)=\frac{1}{\sqrt{\sigma_Y^2(x;\phi)}}f_Z\left[\frac{y-\mu_Y(x;\phi)}{\sqrt{\sigma_Y^2(x;\phi)}}\right].$$

As a consequence, the (conditional) log-likelihood can be written as

$$\ell_p(\phi,\delta,\varkappa,\lambda;Y_t,X_t)=\ell[\mu_Y(X_t;\phi),\sigma_Y^2(X_t;\phi),\delta,\varkappa,\lambda;Y_t]$$

the subscript p is for "parametric" and ℓ was defined in the previous section. Accordingly, we denote the likelihood after reparametrization as $l_p(\phi,\delta,\kappa,\pi;Y_t,X_t)$.

For $\mathcal{P}_a$ part, we only need to check the argument in Step 1 since Steps 2 to 4 are the same. First, notice that for every vector $\boldsymbol{k}$ –with the same dimension as ϕ– such that $|\boldsymbol{k}|=1$ and $(k_2,k_3)\in\{(k,0):k\le 4\}\cup\{(0,k):k\le 2\}\cup\{(1,1)\}$,

$$l_p^{[\boldsymbol{k}_1,k_2,k_3]}(\rho_0)=l_c^{[1,0,k_2,k_3]}(\rho_0)\frac{\partial\mu_Y(X_t;\phi)}{\partial\phi^{\boldsymbol{k}}}+l_c^{[0,1,k_2,k_3]}(\rho_0)\frac{\partial\sigma_Y^2(X_t;\phi)}{\partial\phi^{\boldsymbol{k}}}.$$

Therefore, by the law of iterated expectations, we will have

$$\begin{aligned} E[l_p^{[\boldsymbol{k}_1,k_2,k_3]}(\rho_0)] &= E\{E[l_p^{[\boldsymbol{k}_1,k_2,k_3]}(\rho_0)|X_t]\} \\ &= E\left\{\frac{\partial \mu_Y(X_t;\phi)}{\partial \phi^{\boldsymbol{k}}} E[l_c^{[1,0,k_2,k_3]}(\rho_0)|X_t]\right\} + E\left\{\frac{\partial \sigma_Y^2(X_t;\phi)}{\partial \phi^{\boldsymbol{k}}} E[l_c^{[0,1,k_2,k_3]}(\rho_0)|X_t]\right\} \\ &= 0 \end{aligned}$$

because $E[l_c^{[1,0,k_2,k_3]}(\rho_0)|X_t] = E[l_c^{[0,1,k_2,k_3]}(\rho_0)|X_t] = 0$. Hence, if Assumptions 1 and 2 hold, then the same arguments in Step 1 applies. Analogous arguments apply for the $\mathcal{P}_b$ part too, which completes the proof. □

D Additional lemmas

Lemma 4 *For $k = 1, \ldots, 16$, let*

$$D^k = \Big\{(\eta, \tau, \varphi) : \textit{there exists } \theta \in A^k \textit{ such that } (20)\text{-}(19) \textit{ holds}\Big\}.$$

Then, (i) for all $\theta \in A^k$, there exists a unique $d \in D^k$ such (20)-(19) holds; (ii) for all $d \in D^k$, there exists a unique $\theta \in A^k$ such that (20)-(19) holds.

Proof. (i) is straightforward. As for (ii), we show it for $k = 1$ since the proof for $k = 2, \ldots, 16$ is similar. We only need to show the uniqueness of θ, as the existence follows from the construction of D^1. Note that $\tau > 0$ for all $\theta \in A^1$, thus $\lambda = 1 - \eta/\tau$. With the restrictions of A^1, it holds that

$$\frac{1}{36}\delta^4 - \frac{1}{8}\kappa^2 = \tau, \text{ that is, } \frac{1}{2}\delta\kappa = \varphi\tau. \tag{D1}$$

Hence, we can easily write

$$\frac{2}{9}\delta^4 - \frac{4\tau^2\varphi^2}{\delta^2} = 8\tau. \tag{D2}$$

Since the left hand side of (D2) is strictly increasing in δ^2, we can get unique δ. Finally, we get κ from (D1). □

Lemma 5 *If*

$$\textit{(a) } \sqrt{n}(1-\lambda_n)\delta_n\kappa_n = O_p(1) \quad \textit{and} \quad \textit{(b) } \sqrt{n}(1-\lambda_n)\left[\kappa_n^2 - \frac{2(1-\lambda_n+\lambda_n^2)}{9}\delta_n^4\right] = O_p(1),$$

where $\lambda_n \in [1/2, 1]$, then we have $\sqrt{n}(1-\lambda_n)\kappa_n^2 = O_p(1)$ and $\sqrt{n}(1-\lambda_n)\delta_n^4 = O_p(1)$.

Proof. From (b) we have

$$\sqrt{n}(1-\lambda_n)\kappa_n^2 = \frac{2}{9}(1-\lambda_n+\lambda_n^2)\sqrt{n}(1-\lambda_n)\delta_n^4 + O_p(1).$$

But if $\sqrt{n}(1-\lambda_n)\delta_n^4 = O_p(1)$, then we can trivially show that $\sqrt{n}(1-\lambda_n)\kappa_n^2 = O_p(1)$ because $1 - \lambda_n + \lambda_n^2 \in [3/4, 1]$. The rest of the proof is by contradiction.

Let us assume that $\sqrt{n}(1-\lambda_n)\delta_n^4 \neq O_p(1)$; in other words, that there exists an $\epsilon > 0$ such that for all M_1,

$$\Pr(n^{\frac{1}{2}}(1-\lambda_n)\delta_n^4 > M_1) > \epsilon \text{ i.o.} \tag{D3}$$

Next, given that $\sqrt{n}(1-\lambda_n)\kappa_n^2 - \frac{2}{9}(1-\lambda_n+\lambda_n^2)\sqrt{n}(1-\lambda_n)\delta_n^4 = O_p(1)$, there exists an M_2 such that

$$\Pr\left(\left|\sqrt{n}(1-\lambda_n)\kappa_n^2 - \frac{2}{9}(1-\lambda_n+\lambda_n^2)\sqrt{n}(1-\lambda_n)\delta_n^4\right| < M_2\right) > 1-\frac{\epsilon}{2}$$

for all n. Consider $M' > \max\{M_2, \bar{\delta}^2/6\}$ and let $M_1 = 6M' + 6M_2$. In view of (D3), we have that

$$\Pr[n^{\frac{1}{2}}(1-\lambda_n)\delta_n^4 > 6M' + 6M_2] > \epsilon \quad \text{i.o.}$$

Let

$$A_n = \{n^{\frac{1}{2}}(1-\lambda_n)\delta_n^4 > 6M' + 6M_2\}$$

and

$$B_n = \{|\sqrt{n}(1-\lambda_n)\kappa_n^2 - \frac{2}{9}(1-\lambda_n+\lambda_n^2)\sqrt{n}(1-\lambda_n)\delta_n^4| < M_2\}.$$

Since $\Pr(A_n) > \epsilon$ i.o. and $\Pr(B_n) > 1-\epsilon/2$ for all n, we will also have

$$\Pr(A_n \cap B_n) \geq \Pr(A_n) + \Pr(B_n) - 1 > \frac{\epsilon}{2} \quad \text{i.o.}$$

On the set $A_n \cap B_n$, we have

$$\begin{aligned}
n(1-\lambda_n)^2\delta_n^2\kappa_n^2 &= \sqrt{n}(1-\lambda_n)\delta_n^2\left\{\frac{2}{9}(1-\lambda_n+\lambda_n^2)\sqrt{n}(1-\lambda_n)\delta_n^4\right. \\
&\quad \left. + \left[\sqrt{n}(1-\lambda_n)\kappa_n^2 - \frac{2}{9}(1-\lambda_n+\lambda_n^2)\sqrt{n}(1-\lambda_n)\delta_n^4\right]\right\} \\
&> \sqrt{n}(1-\lambda_n)\delta_n^2\left[\frac{2}{9}(1-\lambda_n+\lambda_n^2)\sqrt{n}(1-\lambda_n)\delta_n^4 - M_2\right] && \text{(D4)} \\
&\geq \sqrt{n}(1-\lambda_n)\delta_n^2\left[\frac{1}{6}\sqrt{n}(1-\lambda_n)\delta_n^4 - M_2\right] && \text{(D5)} \\
&\geq \sqrt{n}(1-\lambda_n)\delta_n^4\frac{M'}{\bar{\delta}^2} && \text{(D6)} \\
&\geq \frac{\sqrt{n}(1-\lambda_n)\delta_n^4}{6} \geq M' + M_2 > M', && \text{(D7)}
\end{aligned}$$

where (D4) uses the definition of B_n, (D5) uses $1-\lambda_n+\lambda_n^2 \geq 3/4$, (D6) combines the definition of A_n with $\delta_n^2 \leq \bar{\delta}^2$, and (D7) uses the definitions of M' and A_n. Hence, $A_n \cap B_n \subset \{n(1-\lambda_n)^2\delta_n^2\kappa_n^2 > M'\}$, which implies that for all M',

$$\Pr[n(1-\lambda_n)^2\delta^2\kappa^2 > M'] \geq \frac{\epsilon}{2} \quad \text{i.o.}$$

which is a contradiction to (a). Thus, we have proved that $\sqrt{n}(1-\lambda_n)\kappa_n^2 = O_p(1)$ and $\sqrt{n}(1-\lambda_n)\delta_n^4 = O_p(1)$, as desired. □

Lemma 6 *(uniform convergence) Denote* $l_0(\theta) = E[l(\theta)]$. *Assume the data is iid,* $E\left(y^2\right) < \infty$ *and* Θ *is compact. Then,*

$$sup_{\theta\in\Theta}\left|\frac{1}{n}L_n(\theta) - l_0(\theta)\right| \xrightarrow{p} 0.$$

Proof. Let $\bar{\sigma}^2 = exp(\bar{\varkappa})/\underline{\lambda} = 2\exp(\bar{\varkappa})$ be an upper bound for $\max(\sigma_1^{*2}, \sigma_2^{*2})$, $\underline{\sigma}^2 = e^{-2\bar{\varkappa}}/(1+\bar{\delta}^2/4)$ a lower bound for $\min(\sigma_1^{*2}, \sigma_2^{*2})$, and $\bar{\mu} = \bar{\delta}$ an upper bound for both $|\mu_1^*|$ and $|\mu_2^*|$. Then, we have

$$\begin{aligned}
l(\theta) &= \log\left\{\lambda\frac{1}{\sqrt{\sigma_1^{*2}}}\exp\left[-\frac{(y-\mu_1^*)^2}{2\sigma_1^{*2}}\right] + (1-\lambda)\frac{1}{\sqrt{\sigma_2^{*2}}}\exp\left[-\frac{y-\mu_2^*)^2}{2\sigma_2^{*2}}\right]\right\} \\
&\geq \lambda\log\left\{\frac{1}{\sqrt{\sigma_1^{*2}}}\exp\left[-\frac{(y-\mu_1^*)^2}{2\sigma_1^{*2}}\right]\right\} + (1-\lambda)\log\left\{\frac{1}{\sqrt{\sigma_2^{*2}}}\exp\left[-\frac{(y-\mu_2^*)^2}{2\sigma_2^{*2}}\right]\right\} \\
&\geq -\frac{1}{2}\log(\bar{\sigma}^2) - \frac{\lambda(y-\mu_1^*)^2 + (1-\lambda)(y-\mu_2^*)^2}{2\underline{\sigma}^2} \\
&\geq -\frac{1}{2}\log(\bar{\sigma}^2) - \frac{(|x|+\bar{\mu})^2}{2\underline{\sigma}^2},
\end{aligned}$$

where the first inequality follows from the concavity of the logarithm, the second one from the definitions of $\bar{\sigma}^2$ and $\underline{\sigma}^2$, and the last one from the definition of $\bar{\mu}$. Moreover,

$$\begin{aligned}
l(\theta) &= \log\left\{\lambda\frac{1}{\sqrt{\sigma_1^{*2}}}\exp\left[-\frac{(y-\mu_1^*)^2}{2\sigma_1^{*2}}\right] + (1-\lambda)\frac{1}{\sqrt{\sigma_2^{*2}}}\exp\left[-\frac{(y-\mu_2^*)^2}{2\sigma_2^{*2}}\right]\right\} \\
&\leq \log\left[\lambda\frac{1}{\sqrt{\sigma_1^{*2}}} + (1-\lambda)\frac{1}{\sqrt{\sigma_2^{*2}}}\right] = \log\left(\frac{1}{\sqrt{\underline{\sigma}^2}}\right).
\end{aligned}$$

Next, letting

$$d(y) = \frac{(|y|+\bar{\mu})^2}{2\underline{\sigma}^2} + \left|\log(\bar{\sigma^2})\right| + \left|\log\left(\frac{1}{\sqrt{\underline{\sigma}^2}}\right)\right|,$$

it is straightforward to see that $|l(\theta)| \leq d(y)$ and $E[|d(y)|] < \infty$. Note that $L_n(\theta)$ is continuous at $\forall\theta\in\Theta$ with probability 1. Thus, by Lemma 2.4 in Newey and McFadden (1994),

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}L_n(\theta) - l_0(\theta)\right| \xrightarrow{p} 0,$$

as desired. □

Lemma 7 *If there exist an* $M_1 > 0$ *and a* $\xi < 1$ *such that* $|H_{3,n}/\sqrt{n}| < M_1$, $|H_{4,n}/\sqrt{n}| < M_1$, $|w_1| > M_1/\xi$, $|w_1| > |w_2|$, $r_n(\theta)/w_1^2 < \xi$, *then* $LR_n(\theta) < 0$.

Proof. We have that

$$LR_n(\theta) = 2\frac{H_{3,n}}{\sqrt{n}}w_1 + 2\frac{H_{4,n}}{\sqrt{n}}w_2 - V_3w_1^2 - V_4w_2^2 + r_n(\theta),$$

so that

$$\begin{aligned}\frac{LR_n(\theta)}{w_1^2} &= 2\frac{H_{3,n}}{\sqrt{n}}\frac{1}{w_1} + 2\frac{H_{4,n}}{\sqrt{n}}\frac{w_2}{w_1^2} - V_3 - V_4\frac{w_2^2}{w_1^2} + \frac{r_n(\delta,\kappa,\lambda)}{w_1^2} \\ &\leq 2\xi + 2\xi\frac{w_2}{w_1} - V_3 + \xi \\ &\leq 5\xi - V_3 \\ &< 0\end{aligned}$$

because $V_3 = E[h_3^2] = 6$, which proves the result. □

Lemma 8 *(Weak convergence)*

(8.1) $\sqrt{n}\left(\frac{1}{n}\tau^{-1}\frac{\partial^2 L(\delta(\tau,\varphi),\kappa(\tau,\varphi),1)}{\partial\lambda^2} - E\left[\tau^{-1}\frac{\partial^2 l(\delta(\tau,\varphi),\kappa(\tau,\varphi),1)}{\partial\lambda^2}\right]\right) = O_{p,(\tau,\varphi)}(1)$.

(8.2) $\sqrt{n}\left(\frac{1}{n}\tau^{-1}\frac{\partial^3 L(\delta(\tau,\varphi),\kappa(\tau,\varphi),\lambda(\eta,\tau,\varphi))}{\partial\lambda^3} - E\left[\tau^{-1}\frac{\partial^3 l(\delta(\tau,\varphi),\kappa(\tau,\varphi),\lambda(\eta,\tau,\varphi))}{\partial\lambda^3}\right]\right) = O_{p,(\tau,\varphi)}(1)$.

(8.3) $\frac{1}{\sqrt{n}}\frac{\partial^6 L_n(\delta,\kappa,1)}{\partial\lambda\partial\delta^i\partial\kappa^j} \Rightarrow \mathcal{G}^{[i,j]}(\delta,\kappa)$ *for* $i+j=5$.

(8.4) $\frac{1}{n}\tau^{-1}\frac{\partial^4 L(\delta(\tau,\varphi),\kappa(\tau,\varphi),\lambda(\eta,\tau,\varphi))}{\partial\lambda^4} = O_{p,(\tau,\varphi)}(1)$.

(8.5) $\tau^{-2}E\left[\frac{\partial^3 l(\delta(\tau,\varphi),\kappa(\tau,\varphi),\lambda(\eta,\tau,\varphi))}{\partial\lambda^3}\right] = O_{(\tau,\varphi)}(1)$.

(8.6) With μ *and* σ^2, $\frac{1}{\sqrt{n}}\frac{\partial^2 L(\rho_{n0})}{\partial\lambda\partial\mu} = O_p(1)$ *and* $\frac{1}{\sqrt{n}}\frac{\partial^2 L(\rho_{n0})}{\partial\lambda\partial\sigma^2} = O_p(1)$.

(8.7) With μ *and* σ^2, $\left\{\frac{1}{n}\tau_n^{-1}\frac{\partial^3 L(\tilde{\rho}_n)}{\partial\lambda^2\partial\mu}\right\} = O_p(1)$ *and* $\left\{\frac{1}{n}\tau_n^{-1}\frac{\partial^3 L(\tilde{\rho}_n)}{\partial\lambda^2\partial\sigma^2}\right\} = O_p(1)$.

Proof. The proofs of (8.1) and (8.2) are similar to the proof of Proposition 1. Therefore, we only give the Taylor expansion of $\partial^2 l(\delta,\kappa,1)/\partial\lambda^2$ and $\partial^3 l(\delta,\kappa,1)/\partial\lambda^3$ to justify the normalization τ^{-1}, but omit the detailed steps. Specifically, a fifth-order Taylor expansions yield

$$\begin{aligned}\frac{\partial^2 l(\delta,\kappa,1)}{\partial\lambda^2} =& h^4\left(\frac{1}{9}\delta^4 - \frac{1}{4}\kappa^2\right) + h_3\delta\kappa \\ &+ \sum_{i=3}^{4}\frac{1}{i!}\frac{\partial^{2+i}l(\delta,\kappa,1)}{\partial\lambda^2\partial\delta^i}\delta^i + \sum_{i+j=3,i\geq1,j\geq1}^{4}\frac{1}{i!j!}\frac{\partial^{2+i+j}l(\delta,\kappa,1)}{\partial\lambda^2\partial\delta^i\partial\kappa^j}\delta^i\kappa^j \\ &+ \sum_{i+j=5}\frac{1}{i!j!}\frac{\partial^{2+i+j}l(\tilde{\delta},\tilde{\kappa},1)}{\partial\lambda^2\partial\delta^i\partial\kappa^j}\delta^i\kappa^j\end{aligned}$$

and

$$\begin{aligned}\frac{\partial^3 l(\delta,\kappa,1)}{\partial\lambda^3} =& 8h^4\delta^4 + \sum_{i=3}^{4}\frac{1}{i!}\frac{\partial^{3+i}l(\delta,\kappa,1)}{\partial\lambda^3\partial\delta^i}\delta^i + \sum_{i+j=3,i\geq1,j\geq1}^{4}\frac{1}{i!j!}\frac{\partial^{3+i+j}l(\delta,\kappa,1)}{\partial\lambda^3\partial\delta^i\partial\kappa^j}\delta^i\kappa^j \\ &+ \sum_{i+j=5}\frac{1}{i!j!}\frac{\partial^{3+i+j}l(\tilde{\delta},\tilde{\kappa},1)}{\partial\lambda^3\partial\delta^i\partial\kappa^j}\delta^i\kappa^j.\end{aligned}$$

The proof of (8.3) is similar but much simpler, as it is not normalized by τ. To prove (8.4), it suffices to apply the uniform law of large numbers (see Lemma 2.4 of Newey and McFadden

(1994)) and use

$$g(\tau,\varphi)=\begin{cases}\tau^{-1}\frac{\partial^4 l(\delta(\tau,\varphi),\kappa(\tau,\varphi),\lambda(\eta,\tau,\varphi))}{\partial\lambda^4} & \text{if } \tau\neq 0,\\ \lim_{\tau\to 0}\tau^{-1}\frac{\partial^4 l(\delta(\tau,\varphi),\kappa(\tau,\varphi),\lambda(\eta,\tau,\varphi))}{\partial\lambda^4}=24h^4 & \text{if } \tau=0.\end{cases}$$

To see (8.5), notice that

$$E\left[\frac{\partial^3 l}{\partial\lambda^3}\right]=-8960\delta^8-54\kappa^4-36\delta^2\kappa^2+o(\tau^2).$$

As for (8.7), we can also show that evaluated at $\tilde{\rho}$,

$$\frac{1}{n}\frac{\partial^3 L_n}{\partial\lambda^2\partial\mu}=-\frac{32}{3\hat{\sigma}}\delta^4\hat{H}_3+\frac{2}{\hat{\sigma}}\kappa^2\hat{H}_3+o_p(\tau)$$

and

$$\frac{1}{n}\frac{\partial^3 L_n}{\partial\lambda^2\partial\sigma^2}=-\frac{16}{3\sigma^2}\frac{1}{n}\hat{H}_4\delta^4+\frac{1}{\sigma^2}\frac{1}{n}\hat{H}_4\kappa^2-\frac{3}{2\sigma^2}\frac{1}{n}\hat{H}_3+o_p(\tau),$$

where

$$\hat{H}_3=\sum_i\hat{y}_i(\hat{y}_i^2-3)\ \text{ and }\ \hat{H}_4=\sum_i\hat{y}_i^4-6\hat{y}_i^2+3\ \text{ with }\ \hat{y}_i=\sum_i\frac{y_i-\hat{\mu}}{\hat{\sigma}},$$

whence we prove the desired result. □

Lemma 9 *$\left|\frac{1}{36}\delta^4-\frac{1}{8}\kappa^2\right|\to 0$ and $\left|\frac{1}{2}\delta\kappa\right|\to 0$ implies $\delta\to 0$ and $\kappa\to 0$.*

Proof. Once again, we prove this by contradiction. If the lemma does not hold, then one of the following statement must be true:
(i) there exist sequences δ_n, κ_n such that $\left|\frac{1}{36}\delta_n^4-\frac{1}{8}\kappa_n^2\right|\to 0$ and $\left|\frac{1}{2}\delta_n\kappa_n\right|\to 0$ but $\delta_n\to\delta^*\neq 0$, or
(ii) there exist sequences δ_n, κ_n such that $\left|\frac{1}{36}\delta_n^4-\frac{1}{8}\kappa_n^2\right|\to 0$ and $\left|\frac{1}{2}\delta_n\kappa_n\right|\to 0$ but $\kappa_n\to\kappa^*\neq 0$.

Consider (i): $\left|\frac{1}{2}\delta_n\kappa_n\right|\to 0$ and $\delta_n\to\delta^*\neq 0$ implies $\kappa_n\to 0$, thus

$$\left|\frac{1}{36}\delta_n^4-\frac{1}{8}\kappa_n^2\right|\to\left|\frac{1}{36}\delta_n^{*4}\right|\neq 0,$$

which is a contradiction to $\left|\frac{1}{36}\delta_n^4-\frac{1}{8}\kappa_n^2\right|\to 0$. Similarly, for (ii), $\left|\frac{1}{2}\delta_n\kappa_n\right|\to 0$ and $\kappa_n\to\kappa^*\neq 0$ implies $\delta_n\to 0$, thus

$$\left|\frac{1}{36}\delta_n^4-\frac{1}{8}\kappa_n^2\right|\to\left|\frac{1}{8}\kappa_n^{*2}\right|\neq 0,$$

as desired. □

References

Amengual, D., Bei, X. and Sentana, E. (2023): "Hypothesis tests with a repeatedly singular information matrix", CEMFI Working Paper 2002, revised May 2023.

Andrews, D.W.K. (1993): "Tests for parameter instability and structural change with unknown change point", *Econometrica* 61, 821–856.

Rotnitzky, A., Cox, D.R., Bottai, M. and Robins, J. (2000): "Likelihood-based inference with singular information matrix", *Bernoulli* 6, 243–284.