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

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 participants 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) and the Women in Econometrics Conference (Toronto, October 2022) for useful comments and suggestions. 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 models play an important role in economics, where they are often used to model unobserved heterogeneity, especially in labor and industrial organization (see Berry, Carnall and Spiller (2006), Cameron and Heckman (1998), and Keane and Wolpin (1997)) but also 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. Mixtures also arise in game theory with multiple equilibria, in measurement error models, as well as in duration models (see Compiani and Kitamura (2016) and the references therein).

In this paper, we focus on finite Gaussian mixtures, which are the most popular. Suppose that individuals can be of two types, $j = 1, 2$ with normal distribution $N(\mu_j, \sigma_j)$ for type j. Assume moreover that the types are not observed by the econometrician. Then, 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. The object of the paper is to test the null hypothesis of a normal versus a finite mixture of two normals.

Testing for normal mixtures is particularly challenging. 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 the null hypothesis, although their identity depends on the way in which one approaches the null. Moreover, when testing $\lambda(1-\lambda) = 0$, λ is on the boundary of the parameter space and standard asymptotic theory fails (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) test requires an eigth-order expansion. All these aspects make testing for normal mixtures highly nonstandard.

Previous papers investigating the properties of the LR tests for normal mixtures include Ghosh and Sen (1985), Hathaway (1985), Chen and Chen (2001), Chen, Chen and Kalbfleisch (2004), Cho and White (2007), and Chen, Ponomareva and Tamer (2014). The closest paper to ours is 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 parameter space under the null 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 moments 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 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 of interest: equality of means and variances, equality of means only, and equality of variances only. In all three cases, our tests are asymptotically equivalent to the analogous LR tests while being much simpler to implement because the unknown mean and variance parameters are estimated under the null hypothesis. Interestingly, when testing for the equality of means and variances, our test boils down to the popular Jarque and Bera's test based on skewness and kurtosis, which implies that theirs 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 normality of the standardized residuals of an econometric model. For that reason, we investigate if our testing procedure is robust to parameter uncertainty. 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 constrained 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 detailed proofs collected in an appendix.

[1] Bai and Ng (2001) propose a test for conditional symmetry in time series contexts based on the empirical distribution function, which can also be used to test the null of normality.

2 Model, hypotheses, and overview of the test

The model we consider is

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

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

Assuming that $\mu(x_i, \alpha) = 0$ and $\sigma(x_i, \alpha) = 1$ without loss of generality, we want to test:
$H_0 : y$ has density $\phi(y_i)$ 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 being unknown parameters. This parametrization guarantees that the marginal distribution of y has zero-mean and unit-variance regardless of the values of the shape parameters. As the labels of the two regimes are not identified, in what follows we set $\lambda \geq 1/2$.[2]

Let $\vartheta = (\delta, \varkappa, \lambda)$, with $\vartheta \in [-\bar{\delta}, \bar{\delta}] \times [-\overline{\varkappa}, \overline{\varkappa}] \times [1/2, 1]$. We consider three different parameter spaces

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

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 mean is common to both regimes.

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

It is well known that the information matrix of the maximum likelihood estimators of $(\delta, \varkappa)$ is singular under H_0. To isolate the singularities and have the first-order derivatives exactly equal to zero under the null, we introduce the following reparametrization:

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

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

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

The goal of our paper is to construct a score-type test for each of the three hypotheses that is asymptotically equivalent to the analogous LR statistic

$$LR_j = 2\left[\sup_{\theta \in \Theta_j} L_n(\delta, \kappa, \lambda) - L_n(\delta, \kappa, 1)\right] \;\; \text{with} \;\; L_n(\delta, \kappa, \lambda) = \sum_{i=1}^{n} l_i(\delta, \kappa, \lambda), \tag{4}$$

where l_i is the log-likelihood of y_i given θ. The main difficulty of finding a score-type test is that some elements of θ are not identified under the null. Indeed, under $H_0 : \delta = \kappa = 0$, the parameter λ is not identified. Similarly, the parameters δ and κ are not identified when $\lambda = 1$. The existing literature circumvents the problem by testing

$$H_{01} : (\delta, \kappa) = 0 \text{ with } \lambda \leq 1 - \varepsilon < 1, \text{ or testing } H_{02} : \lambda = 1 \text{ with } \max\{|\delta|, |\kappa|\} \geq \varepsilon$$

(see, e.g., Cho and White (2007), and Kasahara and Shimotsu (2015), among others). However, the "corner case" $\{(\delta, \kappa, \lambda) : \max\{|\delta|, |\kappa|\} \leq \varepsilon, \lambda \geq 1 - \varepsilon\}$ is missing, and it is not obvious that the resulting test statistic is asymptotically equivalent to (4).

To address this issue, we partition the parameter space as follows,

$$P_{a,j} = \{(\delta, \kappa, \lambda) \in \Theta_j : \max\{|\delta|, |\kappa|\} \leq 1 - \lambda\} \text{ and } P_{b,j} = \{(\delta, \kappa, \lambda) \in \Theta_j : \max\{|\delta|, |\kappa|\} \geq 1 - \lambda\}$$

for $j = 1, 2, 3$ so that we can test the two null hypotheses simultaneously. To the best of our knowledge, this has never been done before.

In what follows, we call $H_{0a,j} : \delta = \kappa = 0$ with $\theta \in P_{a,j}$, and $H_{0b,j} : \lambda = 1$ with $\theta \in P_{b,j}$. To begin with, we treat $H_{0a,j}$ and $H_{0b,j}$ separately and develop the two corresponding test statistics, but then we will combine them by taking the largest of the two.

Specifically, we show in the next section that

$$2\left[\sup_{\theta\in P_{a,1}} L_n(\theta)-L_n(\delta,\kappa,1)\right]=\frac{H_{3,n}^2}{nV_3}+\frac{H_{4,n}^2}{nV_4}+o_p(1),\ \text{ and}$$
$$2\left[\sup_{\theta\in P_{b,1}} L_n(\theta)-L_n(\delta,\kappa,1)\right]=\sup_{(\delta,\kappa):(\delta,\kappa,1)\in\Theta_1\backslash\{0,0,1\}}\left\lfloor\frac{\mathcal{G}_n(\delta,\kappa)}{\sqrt{V(\delta,\kappa)}}\right\rfloor_-^2+o_p(1)$$

where $\lfloor\cdot\rfloor_-=\min(0,\cdot)$,

$$H_{3,n}=\sum_{i=1}^n h_{3i}=\sum_{i=1}^n y_i(y_i^2-3),\qquad V_3=var\,(h_{3,i})=6,$$
$$H_{4,n}=\sum_{i=1}^n h_{4i}=\sum_{i=1}^n(3-6y_i^2+y_i^4)\qquad V_4=var\,(h_{4,i})=24,$$
$$\mathcal{G}_n(\delta,\kappa)=\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{1}{2}(3-e^{\kappa-\frac{\delta^2}{3}})-\frac{1}{\sqrt{e^{\varkappa}}}\exp\left\{\frac{1}{2}\left[y_i^2-(y_i+\delta)^2e^{-\left(\kappa-\frac{\delta^2}{3}\right)}\right]\right\}\right.$$
$$\left.-\delta y_i-(1-e^{\kappa-\frac{\delta^2}{3}})\frac{y_i^2}{2}+\frac{\delta^2}{2}(y_i^2-1)\right],\ \text{ and}$$
$$V(\delta,\kappa)=\frac{\exp[\delta^2/(2-e^{\kappa-\frac{\delta^2}{3}})]}{\sigma^2\sqrt{(2-e^{\kappa-\frac{\delta^2}{3}})e^{\kappa-\frac{\delta^2}{3}}}}-\frac{1}{2\sigma^2}\left[3-2e^{\kappa-\frac{\delta^2}{3}}+(e^{\kappa-\frac{\delta^2}{3}}+\delta^2)^2\right].$$

Thus, we obtain three score-type tests asymptotically equivalent to the respective LR tests, namely,

$$\begin{aligned}LM_1&=\max\left\{\frac{H_{3,n}^2}{nV_3}+\frac{H_{4,n}^2}{nV_4},\sup_{(\delta,\kappa):(\delta,\kappa,1)\in\Theta_1\backslash\{0,0,1\}}\left\lfloor\frac{\mathcal{G}_n(\delta,\kappa)}{\sqrt{V(\delta,\kappa)}}\right\rfloor_-^2\right\}\\&=\sup_{(\delta,\kappa):(\delta,\kappa,1)\in\Theta_1\backslash\{0,0,1\}}\left\lfloor\frac{\mathcal{G}_n(\delta,\kappa)}{\sqrt{V(\delta,\kappa)}}\right\rfloor_-^2,\end{aligned}$$

$$LM_2=\max\left\{\frac{H_{3,n}^2}{nV_3}+\frac{H_{4,n}^2}{nV_4}\mathbf{1}\left[H_{4,n}<0\right],\sup_{|\delta|\le\bar{\delta},|\delta|\ge0}\left\lfloor\frac{\mathcal{G}_n(\delta,\delta^2/3)}{\sqrt{V(\delta,\delta^2/3)}}\right\rfloor_-^2\right\},$$

and

$$LM_3=\max\left\{\frac{H_{4,n}^2}{nV_4}\mathbf{1}\left[H_{4,n}>0\right],\sup_{|\kappa|\le\bar{\varkappa},|\kappa|>0}\left\lfloor\frac{\mathcal{G}_n(0,\kappa)}{\sqrt{V(0,\kappa)}}\right\rfloor_-^2\right\}=\sup_{|\kappa|\le\bar{\varkappa},|\kappa|>0}\left\lfloor\frac{\mathcal{G}_n(0,\kappa)}{\sqrt{V(0,\kappa)}}\right\rfloor_-^2,$$

where the second equalities for LM_1 and LM_3 hold because the test in P_a is no larger than the test in P_b for all possible samples.

3 Test statistics

In this section, we focus on testing whether y_i is a standard normal versus the alternative where y_i is a standardized mixture of two normal distributions. The case with nuisance parameters will be treated in Section 4.

3.1 Test of 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 and one only wants to check that the variances coincide too.

Let $LR_{a,j}$ be the LR statistics for testing $H_{0a,j}$, namely

$$LR_{a,j} = 2\left[\sup_{\theta\in P_{a,j}} L_n(\theta) - L_n(\delta,\kappa,1)\right].$$

Thanks to our reparametrization, the derivatives of the log-likelihood with respect to δ and κ at the point $(0, 0, \lambda)$ are such that

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

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

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

Using an eigth-order expansion of the log-likelihood function, we can characterize the leading terms which are the basis for our score-type tests. In particular, if $\mathbf{1}[A]$ denotes the indicator function for event A, the score-type test statistics corresponding to the three null hypotheses are given by

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

In $LM_{a,1}$ we recognize Jarque and Bera's test, which exploits both the skewness and kurtosis of the data. In contrast, $LM_{a,3}$ exploits only its potential lepkurtosis, while $LM_{a,2}$ both its skewness and its potential platykurtosis. Intuitively, when $\delta = 0$ but $\kappa \neq 0$, the alternative becomes a scale mixture of normals, which can only be leptokurtic and symmetric. On the other hand, when $\kappa = 0$ but $\delta \neq 0$, close to the null we can have either positive or negative skewness but only platykurtosis. Finally, in the unrestricted case there are no restrictions because two-component Gaussian mixtures can generate the entire admissible range of skewness-kurtosis coefficients.

The following propositions establish the equivalence between the LR and our score-type tests and give their asymptotic distributions.

Proposition 1 *For $j = 1, 2, 3$, $LR_{a,j} = LM_{a,j} + o_p(1)$ under $H_{0a,j}$.*

Proposition 2 *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.

3.2 Test of H_{0b}

We are now concerned with testing $H_{0b,j} : \lambda = 1$ with $\theta \in P_{b,j}$. $H_{0b,1}$ 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 and some nuisance parameters are not identified under H_0.

The score with respect to λ at the point $\lambda = 1$ is given by

$$\begin{aligned} \frac{\partial l_i}{\partial \lambda} &= \frac{1}{2}\left(3 - e^{\kappa - \frac{\delta^2}{3}}\right) - \frac{1}{\sqrt{e^{\kappa - \frac{\delta^2}{3}}}} \exp\left\{\frac{1}{2}\left[y_i^2 - (y_i + \delta)^2 e^{-\left(\kappa - \frac{\delta^2}{3}\right)}\right]\right\} \\ &\quad -\delta y_i - \left(1 - e^{\kappa - \frac{\delta^2}{3}}\right)\frac{y_i^2}{2} + \frac{\delta^2}{2}(y_i^2 - 1). \end{aligned} \tag{5}$$

To complicate the analysis further, the score with respect to λ equals zero when δ and κ are 0 simultaneously. For that reason, we first focus on the case where the couple (δ, κ) is away from $(0, 0)$, leaving the discussion of the general case where (δ, κ) may go to $(0, 0)$ for later.

Let $B = \{(\delta, \kappa, 1) \in P_{b,1} : \sqrt{\delta^2 + \kappa^2} \geq \epsilon\}$ for some $\epsilon > 0$. From the form of the score in (5), we can see that the variance of $\partial l_i / \partial \lambda$ becomes unbounded when $e^{-\left(\kappa - \frac{\delta^2}{3}\right)} \geq 2$, so in principle it may seem that we should restrict $\kappa - \delta^2/3 < \ln(2)$. However, our test statistic is based on the ratio of $\partial l_i/\partial\lambda$ over its variance, a ratio that goes to zero for $\kappa - \delta^2/3 \geq \ln(2)$. Therefore, we will never get a maximum in this range because we take the supremum over other values of κ and δ for which the ratio is not 0. As a result, we can ignore this constraint when the variance in the denominator is estimated. In contrast, we need to restrict $\kappa - \delta^2/3 \leq \bar{\varkappa} < \ln(2)$ to avoid numerical overflow when we use the explicit theoretical expression (6) of the variance in computing the test statistic, even though the previous argument still applies.

Lemma 1 *Under H_0, we have*

$$\mathcal{G}_n(\delta,\kappa) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial l_i}{\partial\lambda} \Rightarrow G(\delta,\kappa),$$

where $G(\delta,\kappa)$ is a Gaussian process indexed by $(\delta,\kappa)\in B$ with mean 0, variance given by

$$var[G(\delta,\kappa)] = \frac{\exp[\delta^2/(2-e^{\kappa-\frac{\delta^2}{3}})]}{\sqrt{(2-e^{\kappa-\frac{\delta^2}{3}})e^{\kappa-\frac{\delta^2}{3}}}} - \frac{1}{2}\left[3-2e^{\kappa-\frac{\delta^2}{3}} + (e^{\kappa-\frac{\delta^2}{3}}+\delta^2)^2\right], \tag{6}$$

and covariance $cov[G(\delta_1,\kappa_1), G(\delta_2,\kappa_2)] = g\left(\delta_1, \kappa_1 - \frac{\delta_1^2}{3}, \delta_2, \kappa_2 - \frac{\delta_2^2}{3}\right)$, where

$$\begin{aligned} g(\delta_1,\kappa_1,\delta_2,\kappa_2) &= \frac{\exp\left[-\frac{(\delta_2^2 e^{\kappa_1}+\delta_1^2 e^{\kappa_2})}{2e^{k_1+k_2}}\right]}{\sqrt{e^{\kappa_1}+e^{\kappa_2}-e^{\kappa_1+\kappa_2}}}\exp\left[-\frac{(\delta_2 e^{\kappa_1}+\delta_1 e^{\kappa_2})^2}{2e^{\kappa_1+\kappa_2}(e^{\kappa_1+\kappa_2}-e^{\kappa_1}-e^{\kappa_2})}\right] \\ &\quad -\frac{1}{2}\left[3+2\delta_1\delta_2+(\delta_1\delta_2)^2+\delta_1^2(e^{\kappa_2}-1)+\delta_2^2(e^{\kappa_1}-1)-e^{\kappa_1}-e^{\kappa_2}+e^{\kappa_1+\kappa_2}\right]. \end{aligned} \tag{7}$$

For a given (δ,κ), let

$$LM_b(\delta,\kappa) = \left\lfloor \frac{\mathcal{G}_n(\delta,\kappa)}{\sqrt{var[G(\delta,\kappa)]}} \right\rfloor_-^2.$$

In this context, we can define our test statistic as

$$\widetilde{LM}_{b,1} = \sup_{(\delta,\kappa,1)\in B} LM_b(\delta,\kappa)$$

and the LR test statistic by

$$\widetilde{LR}_{b,j} = 2\left[\sup_{\theta\in P_{b,j}\cap B} L_n(\theta) - L_n(\delta,\kappa,1)\right].$$

We can then show that:

Proposition 3 *Under H_0, we have that*

(a) $$LM_b(\delta,\kappa) \Rightarrow \lfloor \mathcal{G}(\delta,\kappa)\rfloor_-^2,$$

where $\mathcal{G}(\delta,\kappa)$ *is a Gaussian process with zero mean and correlation function*

$$cor[\mathcal{G}(\delta_1,\kappa_1),\mathcal{G}(\delta_2,\kappa_2)] = var[G(\delta_1,\kappa_1)]^{-1/2}cov[G(\delta_1,\kappa_1),G(\delta_2,\kappa_2)]var[G(\delta_2,\kappa_2)]^{-1/2},$$

(b) $$\widetilde{LM}_{b,1} \xrightarrow{d} \sup_{(\delta,\kappa,1)\in B} \lfloor \mathcal{G}(\delta,\kappa)\rfloor_-^2 \text{, } \textit{and}$$

(c) $$\widetilde{LR}_{b,1} = \widetilde{LM}_{b,1} + o_p(1).$$

As we mentioned at the beginning of this section, so far we have restricted (δ,κ) away from 0 for simplicity. But now, we consider the case where $(\delta,\kappa)\to 0$, which is more complex because the score with respect to λ equals zero when δ and κ are simultaneously 0. Consequently,

$$\left\{\frac{\mathcal{G}_n(\delta,\kappa)}{\sqrt{var[G(\delta,\kappa)]}} : (\delta,\kappa,1)\in\Theta_1\backslash\{0,0,1\}\right\}$$

is not Donsker because we could have both $(\delta_{1n},\kappa_{1n})\to 0$ and $(\delta_{2n},\kappa_{2n})\to 0$ but

$$\lim_{n\to\infty}\frac{\mathcal{G}_n(\delta_{1n},\kappa_{1n})}{\sqrt{var[G(\delta_{1n},\kappa_{1n})]}} \neq \lim_{n\to\infty}\frac{\mathcal{G}_n(\delta_{2n},\kappa_{2n})}{\sqrt{var[G(\delta_{2n},\kappa_{2n})]}}.$$

To deal with this problem, we reparametrize the model and define

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

and

$$V'(\tau,m) = \frac{1}{\tau^2}var\{G[\delta(\tau,m),\kappa(\tau,m)]\},$$

so that $\tau\to 0$ if and only if $(\delta,\kappa)\to 0$, in which case $\lim_{\tau\to 0}\mathcal{G}_n'(\tau,m)$ is well defined. We can further show that $\{\mathcal{G}_n'(\tau,m)\}$ is Donsker (see the proof of Proposition 4 for details).

Consider the score-type tests corresponding to $H_{0b,j}$, with $j=1,2,3$, namely:

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

$$LM_{b,2} = \sup_{|\delta|\le\bar{\delta},|\delta|>0} \left\lfloor \frac{\mathcal{G}_n(\delta,\frac{\delta^2}{3})}{\sqrt{var[G(\delta,\frac{\delta^2}{3})]}}\right\rfloor_-^2,$$

and

$$LM_{b,3} = \sup_{|\kappa|\le\bar{\varkappa},|\kappa|>0} \left\lfloor \frac{\mathcal{G}_n(0,\kappa)}{\sqrt{var[G(0,\kappa)]}}\right\rfloor_-^2,$$

where we have excluded the element $\{0,0,1\}$ because at this point $var[G(\cdot)]=0$. Let us explain the choice of the spaces over which the supremum is taken. Recall that

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

When $\lambda = 1$, $(\delta, \kappa, 1) \in \Theta_1$ is equivalent to $\left\{(\delta, \kappa) : |\delta| \le \bar{\delta}, |\kappa - \delta^2/3| \le \bar{\varkappa}\right\}$, i.e.,

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

Similarly, $(\delta, \kappa, 1) \in \Theta_2$ is equivalent to $\left\{(\delta, \kappa) : |\delta| \le \bar{\delta}, \kappa = \delta^2/3\right\}$, 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 our proposed tests and the LR:

Proposition 4 *Under H_0, we have*

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

where

$$LR_{b,j} = 2\left[\sup_{\theta\in P_{b,j}} L_n(\theta) - L_n(\delta,\kappa,1)\right].$$

3.3 Combined test of H_0

Now, we want to test H_0 against H_1 as defined in Section 2. Three tests are available depending on the set Θ_j, $j = 1, 2, 3$, of θ. Note that $\Theta_j = P_{aj} \cup P_{bj}$, so that the likelihood ratio test $LR_j = \max(LR_{a,j}, LR_{b,j})$, and similarly $LM_j = \max(LM_{a,j}, LM_{b,j})$. Using the previous results, we have $LR_j = LM_j + o_p(1)$.

Interestingly, we can show that the test statistic in P_a is no larger than the one in P_b with probability 1 for Θ_1 and Θ_3, which implies that the corresponding tests can be simplified as follows:

Proposition 5 *Under H_0, we have*

$$LR_1 = \sup_{|\delta|\le\bar{\delta}, |\kappa-\delta^2/3|\le\bar{\varkappa}, |\kappa|, |\delta|>0} \left\lfloor \frac{\mathcal{G}_n(\delta,\kappa)}{\sqrt{V(\delta,\kappa)}} \right\rfloor_-^2 + o_p(1) \tag{8}$$

and

$$LR_3 = \sup_{|\kappa|\le\bar{\varkappa}, |\kappa|>0} \left\lfloor \frac{\mathcal{G}_n(0,\kappa)}{\sqrt{V(0,\kappa)}} \right\rfloor_-^2 + o_p(1). \tag{9}$$

However for Θ_2, the test statistic in P_b may be either smaller or larger than that in P_a with positive probability asymptotically (see the appendix for further details).

In summary, our score-type tests are

$$LM_1 = LM_{b,1}, \ \ LM_2 = \max(LM_{a,2}, LM_{b,2}) \ \text{ and } \ LM_3 = LM_{b,3}$$

for testing H_0 against a finite normal mixture with $\theta \in \Theta_1$, Θ_2, and Θ_3, respectively.

3.4 Distribution under local alternatives

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

We first consider local alternatives in which λ goes to 1, namely

$$H_{1n} : \lambda_n = 1 - \frac{\rho}{\sqrt{n}},$$

where ρ is some positive constant and δ and κ are assumed away from 0.

Let P_{β,λ_n}, with $\beta = (\delta, \kappa)$, denote the probability measure of $y_1, \ldots, y_n$ corresponding to H_{1n}, and P_0 be the probability measure of $y_1, \ldots, y_n$ corresponding to H_0. 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 6 *(a) For any $(\beta, 1) \in B$, P_{β,λ_n} is contiguous with respect to P_0.*
(b) Under H_{1n},

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

where

$$c_3 = cov\left(h_{3i}, \frac{\partial l_i}{\partial \lambda}\right) = \delta^3 + 3\delta\left(e^{\kappa - \frac{\delta^2}{3}} - 1\right)$$

and

$$c_4 = cov\left(h_{4i}, \frac{\partial l_i}{\partial \lambda}\right) = 6\delta^2\left(1 - e^{\kappa - \frac{\delta^2}{3}}\right) - \delta^4 - 3\left(1 - e^{\kappa - \frac{\delta^2}{3}}\right)^2.$$

(c) Under H_{1n},

$$LM_{a,1} = \frac{H_{3,n}^2}{nV_3} + \frac{H_{4,n}^2}{nV_4} \xrightarrow{d} \chi^2_2\left(\frac{c_3^2\rho^2}{V_3} + \frac{c_4^2\rho^2}{V_4}\right).$$

(d) Under H_{1n},

$$\widetilde{LM}_{b,1} \xrightarrow{d} \sup_{(\beta,1)\in B} \min\{0, \mathcal{G}(\beta) - var^{1/2}[G(\beta)]\rho\}^2.$$

The following remarks are in order:

1. The $\widetilde{LM}_{b,1}$ test has nontrivial power against local alternatives of order $1/\sqrt{n}$.

2. It follows from Proposition 6 that the LM_1 test has non trivial power against H_{1n} provided either $\delta \neq 0$ or $\kappa \neq 0$. On the other hand, if λ goes to zero faster than $1/\sqrt{n}$, then LM_1 will not have power even if $\delta \neq 0$ and $\kappa \neq 0$.

3. The asymptotic distribution of $\max(LM_{a,1}, \widetilde{LM}_{b,1})$ under H_{1n} could in principle be deduced from Proposition 6 (b), although there is no simple expression for it.

Next, we consider local alternatives in which (δ, κ) approaches $(0, 0)$. Let P_{θ_n} be the distribution of y_i under local alternatives such that $\lim_{n\to\infty}(w_{1n}, w_{2n)} = (w_1, w_2) \in \mathbb{R}^2$, where

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

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

Proposition 7 *(a) P_{θ_n} is contiguous with respect to P_0.*
(b) Under the local alternative P_{θ_n}, we have

$$\begin{aligned} LM_{a,1} &= \frac{H_{3,n}^2}{nV_3} + \frac{H_{4,n}^2}{nV_4} \xrightarrow{d} \chi_2^2\left(V_3 w_1^2 + V_4 w_2^2\right), \\ \widetilde{LM}_{b,1} &= \sup_{(\beta,1)\in B} LM_n(\beta) \xrightarrow{d} \sup_{(\beta,1)\in B} \min\{0, \mathcal{G}(\beta) + var^{-1/2}[G(\beta)](c_3 w_1 + c_4 w_2)\}^2. \end{aligned}$$

An interesting implication of Proposition 7 in terms of power 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 $\widetilde{LM}_{b,1}$ will have nontrivial power against P_{θ_n}. However, if κ_n and δ_n go to zero too fast, or in other words, if $w_1 = o_p(1)$ and $w_2 = o_p(1)$, then the test will have trivial power.

Nevertheless, we would like to emphasize that Propositions 6 and 7 imply 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, and their power goes to 1.

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 an integration constant 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 $k = (k_1, \ldots, k_{d_\theta})$,

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

Proposition 8 *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 8 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 information matrix equality holds.

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 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

$$LR_j = 2\sup_{\theta\in\Theta_j}\sum_{i=1}^{n} L(\hat{y}_i;\theta) - 2\sum_{i=1}^{n} L\left(\hat{y}_i; 0, 0, \frac{1}{2}\right),$$

where the standardized observations are

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

To calculate the maximizers of the first term, we use GlobalSearch Toolbox in Matlab with initial value $(0, 0, 1/2)$ and 1,000 potential starting points for δ and κ. We have also tried as initial values the maximizers of the eighth-order expansion of the log-likelihood function. Specifically, for each Θ_j, we consider:

- 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

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

and

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

In addition, we also tried the optimization of the reparametrized log-likelihood function

$$\sup_{\theta\in\Theta_j} \sum_{i=1}^{n} L(\hat{y}_i;\theta),$$

using analogous initial values. It turns out, though, that the original likelihood with initial value $(0, 0, 1/2)$ yields the value of the parameters that yielded the largest criterion function in all 1,000 trial points.

As for the other tests that we use for comparison purposes, 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), while we fix the Tikhonov regularization parameter α to .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 $y_1, \ldots, y_n$ *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 sup to $|\delta| \leq \bar{\delta} = 2$ and $|\varkappa| \leq \bar{\varkappa} = 2/3$. We then repeat this 100,000 times to get the $1-\alpha$ quantile of the test statistics which we use as critical values.

To assess the size-corrected power of the different tests, we generate $y_1, \ldots, y_n$ from a standardized normal mixture distribution with several combinations of λ, δ and $\varkappa$ that include symmetric mixtures – with either inliers ($\varkappa < 0$) or 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 step 10,000 times. Finally, we compute the corresponding rejection rates using the empirical critical values obtained under the null by means of the parametric bootstrap procedure described in the previous paragraph.

Rejection rates for sample sizes $n = 125$ and $n = 500$ are reported in Tables 1 and 2. We report results for LM_j, $LM_{a,j}$ and $LM_{b,j}$, $j = 1, 2, 3$, whenever different. Note that $LM_{a,1}$ is denoted as JB in the tables because it coincides with Jarque and Bera's test. Moreover, $LM_{b,1}$ and $LM_{b,3}$ are omitted from the tables because LM_1 coincides with $LM_{b,1}$, and LM_3 with $LM_{b,3}$. The upper panels contain results for different combinations of δ and $\varkappa$ when $\lambda = .75$, while the lower ones do the same but when the mixing probability is .95. 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-Bera test, while for Θ_2 (Θ_3) it clearly dominates both LM_a and LM_b (LM_a), as expected. In addition, the relative performance of the tests for different Θ's is in line with the alternative DGPs we consider. Still, the ACS test does a good job, beating both the LR and our score-type tests for some specific alternatives.

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 any given sample are numerically the same regardless of the values of the shape parameters δ, $\varkappa$ and λ, which effectively means that we did not have to re-estimate them under the alternative because they coincide with the sample mean and variance (with denominator n) of the observations. As a result, the criterion function under the alternative calculated keeping μ and σ^2 fixed at their 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 the (log) earnings of individual workers as a function of their education, measured by the number of years of schooling, and their experience, which is usually captured by a quadratic polynomial to reflect skill depreciation. The rationale for these variables is that labor earnings are usually regarded as the returns to human capital, with education and on the job-training being two different forms of investment in it.

The simple Mincer equation, though, fails to capture cross-sectional heterogeneity in the earnings of workers with identical schooling and experience. As an example, it is often argued that female MBAs typically 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.[3] 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.[4]

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.

[3] 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.

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

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 $\hat{\varepsilon}_b$.

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

Importantly, we can achieve higher-order refinements to the asymptotic distribution by imposing the normality of the standardized innovations.

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 recently considered 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. 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 state and prove at the end of the appendix. 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 $N < \infty$.

Proof of Proposition 1

Overview of the proof

In this part, we find the score type test statistic that is asymptotically equivalent to

$$2\left[\sup_{\theta\in P_a} L_n(\theta) - L_n(\delta,\kappa,1)\right],$$

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

$$\begin{aligned} LR_n(\theta) &= 2\left[L_n(\theta) - L_n(0,0,\lambda)\right], \text{ and} \\ LM_n^a(\theta) &= 2\frac{H_{3,n}}{\sqrt{n}}w_1 - V_3 w_1^2 + 2\frac{H_{4,n}}{\sqrt{n}}w_2 - V_4 w_2^2, \end{aligned} \tag{10}$$

where

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

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

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

2. 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} (resp, LM_{a2} and LM_{a3}) in P_{a1} (resp, P_{a2} and P_{a3}).

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

$$LR_n(\theta_n) = LM_n^a(\theta_n) + o_p[h_n(\theta_n)], \tag{11}$$

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

Let l denote the log likelihood of the obervable 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 $(0, 0, \lambda)$ are

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

$$\frac{\partial^2 l}{\partial \delta^2} = 0, \quad \frac{\partial^2 l}{\partial \delta \partial \kappa} = -\frac{1}{2}(1-\lambda)\lambda h_3, \quad \frac{\partial^2 l}{\partial \kappa^2} = \frac{1}{4}(1-\lambda)\lambda h_4,$$
$$\frac{\partial^3 l}{\partial \delta^3} = 0, \quad \text{and} \quad \frac{\partial^4 l}{\partial \delta^4} = -\frac{2}{3}(1-\lambda)\lambda(1-\lambda+\lambda^2)h_4.$$

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

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

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

$$
\sum_{j+k=9} \Delta^{[j,k]} \delta_n^j \kappa_n^k = o_p[h_n(\theta_n)]. \qquad (13)
$$

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

$$
\begin{aligned}
\left|\frac{1}{n}\Delta_n^{[j,k]}\right| \leq& \left|\frac{1}{n}\frac{1}{j!k!} \left.\frac{\partial^{j+k} L_n(\theta)}{\partial\delta^j \partial\kappa^k}\right|_{(0,0,\lambda_n)}\right| + \left|\frac{1}{n}\frac{1}{j!k!} \left.\frac{\partial^{j+k+1} L_n(\theta)}{\partial\delta^{j+1} \partial\kappa^k}\right|_{(\bar{\delta}_n,\bar{\kappa}_n,\lambda_n)}\right| \left|\tilde{\delta}_n\right| \qquad (14) \\
&+ \left|\frac{1}{n}\frac{1}{j!k!} \left.\frac{\partial^{j+k+1} L_n(\theta)}{\partial\delta^j \partial\kappa^{k+1}}\right|_{(\bar{\delta}_n,\bar{\kappa}_n,\lambda_n)}\right| \left|\tilde{\kappa}_n\right| \\
\leq& \left|\frac{1}{j!k!} \left[E \left.\frac{\partial^{j+k} l(\theta)}{\partial\delta^j \partial\kappa^k}\right|_{(0,0,\lambda_n)}\right] + O_p\left(\frac{1}{\sqrt{n}}\right)\right| \\
&+ (1-\lambda_n)\frac{1}{j!k!} \qquad (15) \\
&\times \left\{\left|E\left[\left.\frac{\partial^{j+k+1} l(\theta)}{\partial\delta^{j+1} \partial\kappa^k}\right|_{(0,0,\lambda_n)}\right]\right| + \left|\left[E \left.\frac{\partial^{j+k+1} L_n(\theta)}{\partial\delta^j \partial\kappa^{k+1}}\right|_{(0,0,\lambda_n)}\right]\right| + o_p(1)\right\} \\
=& O\left[(1-\lambda_n)^2\right] + O_p\left(\frac{1}{\sqrt{n}}\right) + o_p(1-\lambda_n), \qquad (16)
\end{aligned}
$$

where (14) is a Taylor expansion around $(0, 0, \lambda_n)$, (15) follows from the central limit theorem

and

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

while (16) 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,

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

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

If we then use (12) and (13), we can show that

$$\begin{aligned}\frac{1}{2}LR_n(\theta_n) =& \sqrt{n}\delta_n^4\left(A_{1n} + \sqrt{n}\delta_n^4 A_{3n}\right) + \sqrt{n}\kappa_n^2\left(A_{4n} + \sqrt{n}\kappa_n^2 A_{6n}\right) \\ &+ \sqrt{n}\delta_n\kappa_n\left(A_{8n} + \sqrt{n}\delta_n\kappa_n A_{13n}\right) + o_p[h_n(\theta_n)],\end{aligned} \tag{17}$$

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,

$$\begin{aligned}\frac{1}{2}LR_n(\theta_n) =& -\frac{\lambda_n(1-\lambda_n+\lambda_n^2)}{36}\frac{H_{4,n}}{\sqrt{n}}\sqrt{n}(1-\lambda_n)\delta_n^4 \\ &-\frac{1}{2}\left[\frac{\lambda_n(1-\lambda_n+\lambda_n^2)}{36}\right]^2 V_4 n(1-\lambda_n)^2\delta_n^8 \\ &+\frac{\lambda_n}{8}\frac{H_{4,n}}{\sqrt{n}}\sqrt{n}(1-\lambda_n)\kappa_n^2 - \frac{1}{2}\left(\frac{\lambda_n}{8}\right)^2 V_4 n(1-\lambda_n)^2\kappa_n^4 \\ &-\frac{\lambda_n}{2}\frac{H_{3,n}}{\sqrt{n}}\sqrt{n}(1-\lambda_n)\delta_n\kappa_n - \frac{1}{2}\left(\frac{\lambda_n}{2}\right)^2 V_3 n(1-\lambda_n)^2\delta_n^2\kappa_n^2 + o_p[h_n(\theta_n)] \qquad (18) \\ =& \frac{H_{3,n}}{\sqrt{n}} w_{1n} - \frac{1}{2}V_3 w_{1n}^2 + \frac{H_{4,n}}{\sqrt{n}} w_{2n} - \frac{1}{2}V_4 w_{2n}^2 + o_p[h_n(\theta_n)], \qquad (19)\end{aligned}$$

with

$$w_{1n} = -\frac{\lambda_n}{2}\sqrt{n}(1-\lambda_n)\delta_n\kappa_n \quad \text{and} \quad w_{2n} = -\frac{\lambda_n(1-\lambda_n+\lambda_n^2)}{36}\sqrt{n}(1-\lambda_n)\delta_n^4 + \frac{\lambda_n}{8}\sqrt{n}(1-\lambda_n)\kappa_n^2, \tag{20}$$

where in the first step we re-write (17) as (18). 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 (18) to (19). 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 we wanted to show.

Step 2

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

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

It is then straightforward to see that $w_{1n}^{LM} = O_p(1)$ and $w_{2n}^{LM} = O_p(1)$, where w_{1n}^{LM} and w_{2n}^{LM} are defined in (20), 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

$$\sqrt{n}(1-\lambda_n^{LM})\delta_n^{LM}\kappa_n^{LM} = O_p(1). \tag{21}$$

In addition, we also have

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

Then by Lemma 7, $\sqrt{n}(1-\lambda_n^{LM})\left(\kappa_n^{LM}\right)^2 = O_p(1)$ and $\sqrt{n}(1-\lambda_n^{LM})\left(\delta_n^{LM}\right)^4 = O_p(1)$. Together with (21), 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

In what follows, 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 8, we have

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}L_n(\theta) - l_0(\theta)\right| \xrightarrow{p} 0 \tag{22}$$

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

$$l_0\left(0,0,\lambda\right) > \sup_{\delta^2+\kappa^2>\epsilon, \theta\in P_a} l_0(\theta) \tag{23}$$

(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 $\Pr(C_n) \geq 1 - \epsilon/4$ 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

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

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

Then, we can show that for all M,

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

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 9 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

$$\begin{aligned} n\left(1-\lambda_n^{LR}\right)^2\left(\kappa_n^{LR}\delta_n^{LR}\right)^2 &= n^{\frac{1}{2}}\left(1-\lambda_n^{LR}\right)\left(\kappa_n^{LR}\right)^2 n^{\frac{1}{2}}\left(1-\lambda_n^{LR}\right)\left(\delta_n^{LR}\right)^2 \\ &= \left\{\frac{8w_{2n}^{LR}}{\lambda_n^{LR}} + \frac{2}{9}[1-\lambda_n^{LR} + \left(\lambda_n^{LR}\right)^2]n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\delta_n^{LR}\right)^4\right\} \qquad (24) \\ &\quad \times n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\delta_n^{LR}\right)^2 \\ &\geq \left[-16\left|w_{2n}^{LR}\right| + \frac{1}{6}n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\delta_n^{LR}\right)^4\right] n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\delta_n^{LR}\right)^2 \qquad (25) \\ &\geq \left(\frac{1}{6} - \frac{1}{18}\right) n\left(1-\lambda_n^{LR}\right)^2\left(\delta_n^{LR}\right)^6 \geq \frac{n\left(1-\lambda_n^{LR}\right)^2\left(\delta_n^{LR}\right)^8}{9\xi^2}, \qquad (26) \end{aligned}$$

where (24) follows from the definition of w_{2n}^{LR}, (25) follows from the bound of λ_n^{LR}, the first inequality of (26) 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

$$\left|w_{1n}^{LR}\right| = \frac{\lambda_n^{LR}}{2}\left|n^{\frac{1}{2}}(1-\lambda_n^{LR})\kappa_n^{LR}\delta_n^{LR}\right| \geq \frac{1}{4}\frac{n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\delta_n^{LR}\right)^4}{3\xi} \tag{27}$$

$$\begin{cases} \geq \frac{24M}{\xi} > \frac{M_1}{\xi} & \text{(i)} \\ > \frac{1}{288}n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\delta_n^{LR}\right)^4 \geq |w_{2n}^{LR}| & \text{(ii)} \end{cases} \tag{28}$$

where (27) follows from (26), (28i) follows from combining A_n with E_n and $M_1 < M$, while (28ii) 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

$$\left|n^{\frac{1}{2}}(1-\lambda_n^{LR})\kappa_n^{LR}\delta_n^{LR}\right| \geq \frac{n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\delta_n^{LR}\right)^4}{3\xi} \geq n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\delta^{LR}\right)^4 \tag{29}$$

$$\left|n^{\frac{1}{2}}(1-\lambda_n^{LR})\kappa_n^{LR}\delta_n^{LR}\right| \geq \frac{n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\delta_n^{LR}\right)^4}{3\xi} \geq \frac{2n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\kappa^{LR}\right)^2}{3\xi} > n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\kappa^{LR}\right)^2, \tag{30}$$

where (29) follows from (26) and $\xi < 1/3$, and (30) follows from the definition of E_n and $\xi < 1/3$. Thus, $h_n(\theta_n^{LR}) = n\left(1-\lambda_n^{LR}\right)^2\left(\kappa_n^{LR}\delta_n^{LR}\right)^2$ and, as a result,

$$\begin{aligned} \left|\frac{r_n(\theta_n^{LR})}{\left(w_{1n}^{LR}\right)^2}\right| &= \left|\frac{r_n(\theta_n^{LR})}{h_n(\theta_n^{LR})}\frac{h_n(\theta_n^{LR})}{\left(w_{1n}^{LR}\right)^2}\right| = \left|\frac{r_n(\theta_n^{LR})}{h_n(\theta_n^{LR})}\right|\left|\frac{n\left(1-\lambda_n^{LR}\right)^2\left(\kappa_n^{LR}\delta_n^{LR}\right)^2}{\left(w_{1n}^{LR}\right)^2}\right| \\ &< \xi\left(\frac{1}{288}\right)^2\frac{4}{[\lambda_n^{LR}]^2} < \xi, \end{aligned} \tag{31}$$

where (31) 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 9, 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 9 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

$$
\begin{aligned}
n\left(1-\lambda_n^{LR}\right)^2\left(\kappa_n^{LR}\delta_n^{LR}\right)^2 &\geq n^{\frac{1}{2}}\left(\kappa_n^{LR}\right)^2 n^{\frac{1}{2}}\left(\delta_n^{LR}\right)^4\frac{1}{\xi^2}\left(1-\lambda_n^{LR}\right)^2 &(32)\\
&\geq n^{\frac{1}{2}}\left(1-\lambda_n^{LR}\right)\left(\kappa_n^{LR}\right)^2\frac{36}{(1-\lambda_n+\lambda_n^2)} &(33)\\
&\times\left(\frac{1}{8}\sqrt{n}\left(1-\lambda_n^{LR}\right)\left(\kappa_n^{LR}\right)^2-\frac{w_{2n}^{LR}}{\lambda_n}\right)\frac{1}{\xi^2}\\
&\geq n^{\frac{1}{2}}\left(1-\lambda_n^{LR}\right)\left(\kappa_n^{LR}\right)^2 36 &(34)\\
&\times\left(\frac{1}{8}\sqrt{n}\left(1-\lambda_n^{LR}\right)\left(\kappa_n^{LR}\right)^2-2|w_{2n}^{LR}|\right)\frac{1}{\xi^2}\\
&\geq 4n\left(1-\lambda_n^{LR}\right)^2\left(\kappa_n^{LR}\right)^4\frac{1}{\xi^2}, &(35)
\end{aligned}
$$

where (32) follows from the definition of C_n, (33) follows from the definition of w_{2n}^{LR}, (34) follows from the bound of λ_n^{LR}, and (35) follows from combining D_n with E_n^c.

Then,

$$
\left|w_{1n}^{LR}\right| = \left|\frac{(1-\lambda_n^{LR})\lambda_n^{LR}}{2}n^{\frac{1}{2}}\kappa_n^{LR}\delta_n^{LR}\right| \geq \frac{1}{4}\frac{2n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\kappa_n^{LR}\right)^2}{\xi} > \frac{1}{72}n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\kappa_n^{LR}\right)^2 \quad (36)
$$

$$
\begin{cases} > M > \frac{M_1}{\xi} & \text{(i)} \\ \geq |w_2^{LR}| & \text{(ii)} \end{cases}, \quad (37)
$$

where (36) follows from (35), (37i) follows from combining A_n with E_n^c, and (37ii) 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

$$
\left|n^{\frac{1}{2}}(1-\lambda_n^{LR})\kappa_n^{LR}\delta_n^{LR}\right| \geq \frac{2n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\kappa_n^{LR}\right)^2}{\xi} > n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\kappa_n^{LR}\right)^2 \quad (38)
$$

$$
\begin{aligned}
\left|n^{\frac{1}{2}}(1-\lambda_n^{LR})\kappa_n^{LR}\delta_n^{LR}\right| &\geq \frac{2n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\kappa_n^{LR}\right)^2}{\xi} > \frac{n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\delta_n^{LR}\right)^4}{\xi}\\
&> n^{\frac{1}{2}}(1-\lambda_n^{LR})\left(\delta_n^{LR}\right)^4, \quad (39)
\end{aligned}
$$

where (38) follows from (35), and (39) follows from the definition of E_n^c. Thus, $h_n(\theta_n^{LR}) = n\left(1-\lambda_n^{LR}\right)^2\left(\kappa_n^{LR}\delta_n^{LR}\right)^2$ and, consequently,

$$
\left|\frac{r_n(\theta_n^{LR})}{\left(w_{1n}^{LR}\right)^2}\right| = \left|\frac{r_n(\theta_n^{LR})}{h_n(\theta_n^{LR})}\right|\left|\frac{n\left(1-\lambda_n^{LR}\right)^2\left(\kappa_n^{LR}\delta_n^{LR}\right)^2}{\left(w_{1n}^{LR}\right)^2}\right| = \left|\frac{r_n(\theta_n^{LR})}{h_n(\theta_n^{LR})}\right|\left|\frac{4}{(\lambda_n^{LR})^2}\right| < \xi, \quad (40)
$$

where the last inequality in (40) follows from the definition of C_n. By Lemma 9, 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 9 once again to show that $A_n \cap B_n \cap C_n \cap D_n^c \cap F_n \subset \left\{LR(\theta_n^{LR}) < 0\right\} = \emptyset$. 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,

$$\begin{aligned}
\left|\frac{r_n(\theta_n^{LR})}{\left(w_{1n}^{LR}\right)^2}\right| &= \left|\frac{r_n(\theta_n^{LR})}{h_n(\theta_n^{LR})}\right| \quad (41)\\
&\times \left|\frac{\max\left\{1, n\left(1-\lambda_n^{LR}\right)^2\left(\kappa_n^{LR}\right)^4, n\left(1-\lambda_n^{LR}\right)^2\left(\delta_n^{LR}\right)^8, n\left(1-\lambda_n^{LR}\right)^2\left(\kappa_n^{LR}\delta_n^{LR}\right)^2\right\}}{\left(w_{1n}^{LR}\right)^2}\right|\\
&< \left|\frac{r_n(\theta_n^{LR})}{h_n(\theta_n^{LR})}\right| \left|\frac{\max\left\{\left(288 w_{2n}^{LR}\right)^2, \left(2w_{1n}^{LR}/\lambda_n^{LR}\right)^2\right\}}{\left(w_{1n}^{LR}\right)^2}\right| \quad (42)\\
&\le \left|\frac{r_n(\theta_n^{LR})}{h_n(\theta_n^{LR})}\right| (288)^2 \le \xi,
\end{aligned}$$

where (41) to (42) follow from the definitions of D_n^c and w_1. By Lemma 9, 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

$$\begin{aligned}
\frac{h_n(\theta_n^{LR})}{\left(w_{2n}^{LR}\right)^2} &= \frac{\max\left\{n\left(1-\lambda_n^{LR}\right)^2\left(\kappa_n^{LR}\right)^4, n\left(1-\lambda_n^{LR}\right)^2\left(\delta_n^{LR}\right)^8, n\left(1-\lambda_n^{LR}\right)^2\left(\kappa_n^{LR}\delta_n^{LR}\right)^2\right\}}{\left(w_{2n}^{LR}\right)^2}\\
&\le \frac{\max\left\{\left(288 w_{2n}^{LR}\right)^2, \left(4w_{1n}^{LR}\right)^2\right\}}{\left(w_{2n}^{LR}\right)^2} \le 12^4 \times 4, \quad (43)
\end{aligned}$$

where the first inequality in (43) follows from the definition of D_n^c and the second one from the definition of F_n^c. But then,

$$\begin{aligned}
\frac{LR_n(\theta_n^{LR})}{\left(w_{2n}^{LR}\right)^2} &= 2\frac{H_{3,n}}{\sqrt{n}}\frac{w_{1n}^{LR}}{\left(w_{2n}^{LR}\right)^2} + 2\frac{H_{4,n}}{\sqrt{n}}\frac{1}{w_{2n}^{LR}} - V_3\frac{\left(w_{1n}^{LR}\right)^2}{\left(w_{2n}^{LR}\right)^2} - V_4 + \frac{r_n(\theta_n^{LR})}{\left(w_2^{LR}\right)^2}\\
&\le 2\frac{M_1}{M} + 2\frac{M_1}{M} - V_4 + \left|\frac{r_n(\theta_n^{LR})}{h_n(\theta_n^{LR})}\right| \times 12^4 \times 4 \quad (44)\\
&\le 4\xi - V_4 + \xi < 0, \quad (45)
\end{aligned}$$

where (44) follows from the combination of A_n with B_n, D_n^c, F_n^c and (43), and (45) 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

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

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

Step 3.2.2

Next, we will show that $n\left(1-\lambda_n^{LR}\right)^2\left(\delta_n^{LR}\kappa_n^{LR}\right)^2=O_p(1)$, i.e. that for all $\epsilon>0$, there exists $M>1$ such that $\Pr[n\left(1-\lambda_n^{LR}\right)^2\delta_n^{LR}\kappa_n^{LR2}>M]<\epsilon$ 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 $n\left(1-\lambda_n^{LR}\right)^2\left(\delta_n^{LR}\right)^8=O_p(1)$ and $n\left(1-\lambda_n^{LR}\right)^2\left(\kappa_n^{LR}\right)^4=O_p(1)$. Letting $0<m<\frac{1}{4}V_3$, we have that

$$\Pr\left(\left|\frac{16r_n(\theta_n^{LR})}{\max\{1,n\left(1-\lambda_n^{LR}\right)^2\left(\delta_n^{LR}\kappa_n^{LR}\right)^2\}}\right|>2m\right)<\frac{\epsilon}{2}\quad\text{ult.}\tag{46}$$

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 $\forall n$,

$$\Pr\left[\frac{H_{3,n}}{\sqrt{n}}\geq M\left(\frac{V_3}{2}-2m\right)\right]<\frac{\epsilon}{4}\text{ and }\Pr\left[\frac{\left(\frac{H_{4,n}}{\sqrt{n}}\right)^2}{2V_4}>mM^2\right]<\frac{\epsilon}{4}.\tag{47}$$

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

$$\begin{aligned}
=&\Pr\left[\{\left|w_{1n}^{LR}\right|>M\}\cap\{LR(\theta_n^{LR})\geq 0\}\right]\\
=&\Pr\left[\{\left|w_{1n}^{LR}\right|>M\}\cap\left\{\frac{LM_n^a(\theta_n^{LR})}{(w_{1n}^{LR})^2}+\frac{r_n(\theta_n^{LR})}{(w_{1n}^{LR})^2}\geq 0\right\}\right]\\
=&\Pr\left[\{\left|w_{1n}^{LR}\right|>M\}\cap\left\{\frac{LM_n^a(\theta_n^{LR})}{(w_{1n}^{LR})^2}+\frac{r_n(\theta_n^{LR})}{(w_{1n}^{LR})^2}\geq 0\right\}\cap\left\{\left|\frac{r_n(\theta_n^{LR})}{(w_{1n}^{LR})^2}\right|\leq 2m\right\}\right]\\
&+\Pr\left[\{\left|w_{1n}^{LR}\right|>M\}\cap\left\{\frac{LM_n^a(\theta_n^{LR})}{(w_{1n}^{LR})^2}+\frac{r_n(\theta_n^{LR})}{(w_{1n}^{LR})^2}\geq 0\right\}\cap\left\{\left|\frac{r_n(\theta_n^{LR})}{(w_{1n}^{LR})^2}\right|>2m\right\}\right]\\
\leq&\Pr\left[\{\left|w_{1n}^{LR}\right|>M\}\cap\left\{\frac{H_{3,n}}{\sqrt{n}}\frac{1}{w_{1n}^{LR}}-\frac{V_3}{2}-\frac{V_4\left(w_{2n}^{LR}-\frac{1}{V_4}\frac{H_{4,n}}{\sqrt{n}}\right)^2}{2(w_{1n}^{LR})^2}+\frac{\frac{1}{V_4}\left(\frac{H_{4,n}}{\sqrt{n}}\right)^2}{2(w_{1n}^{LR})^2}+m\geq 0\right\}\right]\\
&+\Pr\left[\left|\frac{r_n(\theta_n^{LR})}{\left(w_{1n}^{LR}\right)^2}\right|>2m\right]\\
\leq&\Pr\left(\{\left|w_{1n}^{LR}\right|>M\}\cap\left\{\frac{H_{3,n}}{\sqrt{n}}\geq w_{1n}^{LR}\left[\frac{V_3}{2}-m-\frac{H_{4,n}^2}{2nV_4}\frac{1}{\left(w_{1n}^{LR}\right)^2}\right]\right\}\right)+\frac{\epsilon}{2} &&(48)\\
\leq&\Pr\left[\frac{H_{3,n}}{\sqrt{n}}\geq M\left(\frac{V_3}{2}-m-\frac{H_{4,n}^2}{2nV_4}\frac{1}{M^2}\right)\right]+\frac{\epsilon}{2}\quad\text{ult.}, &&(49)
\end{aligned}$$

where (48) uses (46). In addition,

$$
\begin{aligned}
(49) \leq & \Pr\left[\left\{\frac{H_{3,n}}{\sqrt{n}} \geq M\left(\frac{V_3}{2} - m - \frac{H_{4,n}^2}{2nV_4}\frac{1}{M^2}\right)\right\} \cap \left\{\frac{H_{4,n}^2}{2nV_4} \leq mM^2\right\}\right] \\
& + \Pr\left[\left\{\frac{H_{3,n}}{\sqrt{n}} \geq M\left(\frac{V_3}{2} - m - \frac{H_{4,n}^2}{2nV_4}\frac{1}{M^2}\right)\right\} \cap \left\{\frac{H_{4,n}^2}{2nV_4} > mM^2\right\}\right] + \frac{\epsilon}{2} \\
\leq & \Pr\left[\frac{H_{3,n}}{\sqrt{n}} \geq M\left(\frac{V_3}{2} - 2m\right)\right] + \Pr\left(\frac{H_{4,n}^2}{2nV_4} > mM^2\right) + \frac{\epsilon}{2} \\
\leq & \frac{\epsilon}{4} + \frac{\epsilon}{4} + \frac{\epsilon}{2} = \epsilon, \qquad (50)
\end{aligned}
$$

where in (50) we have used (47).

Step 4

We now show that $LR_n(\theta_n^{LR}) = LM_n^a(\theta_n^{LM}) + o_p(1)$, that is, that $\forall\epsilon_1 > 0, \forall\epsilon_2 > 0$, 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

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

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

$$\Pr\left(\max G_n \leq M\right) > 1 - \frac{\epsilon_2}{2}. \qquad (51)$$

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

$$\Pr\left(\sup_{\theta\in A}|LR_n(\theta) - LM_n^a(\theta)| < \epsilon_1\right) > 1 - \frac{\epsilon_2}{2}. \qquad (52)$$

To show this, let

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

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 (11). Therefore, for $n > N$ we have

$$\begin{aligned}
&\Pr\left(\left|LR_n(\theta_n^{LR}) - LM_n^a(\theta_n^{LM})\right| < \epsilon_1\right) \\
\geq &\Pr\left(\left\{\left|LR_n(\theta_n^{LR}) - LM_n^a(\theta_n^{LM})\right| < \epsilon_1\right\} \cap \left\{\theta_n^{LR} \in A\right\} \cap \left\{\theta_n^{LM} \in A\right\}\right) \\
\geq &\Pr\left(\left\{\sup_{\theta\in A}|LR_n(\theta) - LM_n^a(\theta)| < \epsilon_1\right\} \cap \left\{\theta_n^{LR} \in A\right\} \cap \left\{\theta_n^{LM} \in A\right\}\right) && (53) \\
\geq &\Pr\left(\sup_{\theta\in A}|LR_n(\theta) - LM_n^a(\theta)| < \epsilon_1\right) + P\left(\left\{\theta_n^{LR} \in A\right\} \cap \left\{\theta_n^{LM} \in A\right\}\right) - 1 && (54) \\
\geq &1 - \frac{\epsilon_2}{2} + 1 - \frac{\epsilon_2}{2} - 1 = 1 - \epsilon_2, && (55)
\end{aligned}$$

where we have used $\Pr(E_1 \cap E_2) \geq \Pr(E_1) + \Pr(E_2) - 1$ to go from (53) to (54), and (51) and (52) to go from (54) to (55).

Step 5

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

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 $w_{21} = \frac{(1-\lambda)\lambda}{8}\sqrt{n}\kappa^2$ and $w_{22} = -\frac{(1-\lambda)\lambda(1-\lambda+\lambda^2)}{36}\sqrt{n}\delta^4$. We first aim to find an upper bound for $LM_n^a(\theta_n^{LM})$. In that respect, we can easily show that

$$LM_n^a(\theta_n^{LM}) \leq \frac{H_{3,n}^2}{nV_3} + \frac{H_{4,n}^2}{nV_4}. \tag{56}$$

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 $(\delta_n^*, \kappa_n^*, \lambda_n^*) \in P_a$ with probability approaching one, whence

$$LM_n^a\left(\theta_n^{LM}\right) \geq LM_n^a\left(\delta_n^*, \kappa_n^*, \lambda_n^*\right) = \frac{H_{3,n}^2}{nV_3} + \frac{H_{4,n}^2}{nV_4} + o_p(1). \tag{57}$$

To verify the second equality of (57), we can easily check by hand that

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

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

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

with

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

But then, (56) and (57) 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 $\bar{\lambda} \in (1/2, 1)$,

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

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

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

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

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 $P_{a,3} = \{(\delta, \kappa, \lambda) : (\delta, \kappa - (2\lambda - 1)\delta^3/3, \lambda) \in \Theta_3', \max\{|\delta|, |\kappa|\} \leq 1 - \lambda\}$. 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

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

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

Proof of Lemma 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) $\tilde{B}$ (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 $\beta \equiv (\delta, \kappa) \in \widetilde{B} = \left\{ (\delta, \kappa) : (\delta, \kappa, 1) \in P_{b,1} \text{ and } \sqrt{\delta^2 + \kappa^2} \geq \epsilon \right\}$ and $\widetilde{B}$ is compact.

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

$$E \sup_{\beta \in \widetilde{B}} \left| \frac{\partial l}{\partial \lambda} (\beta, 1) \right| < \infty. \tag{58}$$

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 $\widetilde{B}$, 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

$$\begin{aligned}
\frac{1}{\sqrt{e^{\varkappa}}}\exp\left[\frac{1}{2}\left\{y^2-\frac{[y+\delta]^2}{e^{\varkappa}}\right\}\right] &= e^{-\varkappa/2}\exp\left[-\frac{1}{2}\frac{(1-e^{\varkappa})}{e^{\varkappa}}y^2\right]\exp\left(-\frac{y\delta}{e^{\varkappa}}\right)\exp\left(-\frac{\delta^2}{2e^{\varkappa}}\right) \\
&= e^{-\varkappa/2}\exp\left[\frac{1}{2}\left(1-e^{-\varkappa}\right)y^2\right]\exp\left(-\frac{y\delta}{e^{\varkappa}}\right)\exp\left(-\frac{\delta^2}{2e^{\varkappa}}\right) \\
&\leq \exp\left[\frac{1}{2}\left(1-e^{-\overline{\varkappa}}\right)y^2\right]\exp\left(\frac{|y|\,|\delta|}{e^{\varkappa}}\right) \\
&\leq \exp\left[\frac{1}{2}\left(1-e^{-\overline{\varkappa}}\right)y^2\right]\exp\left(|y|\,e^{\overline{\varkappa}}\,|\overline{\delta}|\right) \\
&\equiv g^*(y). \qquad (59)
\end{aligned}$$

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 (58), 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 (7).

(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

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

But

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

Using the mean-value theorem, we have

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

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

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

where $\widetilde{\beta} = \left(\tilde{\delta}, \tilde{\kappa}\right)$, $\tilde{\delta}$ is between δ_1 and δ_2, and g^* is defined in (59). 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:

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

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). □

Proof of Proposition 3

Expressions (a) and (b) are direct consequences of Lemma 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, $\sigma_1^* \geq [\left(1 + \overline{\delta}^2\right)(1 + \exp(\overline{\varkappa}))]^{-1} > 0$ and $\sigma_2^* \geq \exp(-\overline{\varkappa})[\left(1 + \overline{\delta}^2\right)(1 + \exp(\overline{\varkappa}))]^{-1} > 0$.

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

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

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 $\beta \in \tilde{B}$. 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^+, $\forall \beta \in \tilde{B}$.

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

Assumption 3^* holds by Lemma 1. 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 notation) corresponds to our λ, and (δ, ψ) (in Andrews notation) is 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 $\widetilde{LR}_{b,1} = \widetilde{LM}_{b,1} + o_p(1)$. □

Proof of Proposition 4

Overview of the proof

In this part, we find the score-type test statistic that is asymptotically equivalent to

$$2 \left[\sup_{\theta \in P_b} L_n(\theta) - L_n(\delta, \kappa, 1) \right]$$

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

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

and for $\theta \in P_b \backslash \{(0,0,1)\}$, let

$$LM_n^b(\theta) = \frac{2}{\sqrt{n}}\frac{\partial L_n(\delta,\kappa,1)}{\partial \lambda}\sqrt{n}(\lambda-1) - V(\delta,\kappa)n(\lambda-1)^2,$$

$$V_b(\delta,\kappa) = E\left[\left(\frac{\partial l(\delta,\kappa,1)}{\partial\lambda}\right)^2\right].$$

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

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

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). One way to circumvent this problem is to normalize $\frac{\partial l(\delta,\kappa,1)}{\partial\lambda}$ by a function of (δ,κ) and further reparameterize the model. To be more specific, for $\delta^2+\kappa^2>0$, let

$$\eta = \max\left\{\left|\frac{1}{36}\delta^4 - \frac{1}{8}\kappa^2\right|, \left|\frac{1}{2}\delta\kappa\right|\right\}(1-\lambda), \tag{60}$$

$$\tau = \max\left\{\left|\frac{1}{36}\delta^4 - \frac{1}{8}\kappa^2\right|, \left|\frac{1}{2}\delta\kappa\right|\right\}, \tag{61}$$

$$m = \frac{\min\left\{\left|\frac{1}{36}\delta^4 - \frac{1}{8}\kappa^2\right|, \left|\frac{1}{2}\delta\kappa\right|\right\}}{\max\left\{\left|\frac{1}{36}\delta^4 - \frac{1}{8}\kappa^2\right|, \left|\frac{1}{2}\delta\kappa\right|\right\}}. \tag{62}$$

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

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

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

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

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

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

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

Define $A_{i1} = P_b \backslash (A_{i0} \cup \{(0,0,1)\})$ 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

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

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

$$D^k = \left\{d = (\eta,\tau,m) : \text{there exists } \theta \in A^k \text{ such that (60)-(62) holds}\right\}.$$

Similarly, let

$$\begin{aligned}
A^k_{\delta\kappa} &= \left\{(\delta,\kappa) : \text{there exists } \lambda \text{ such that } (\delta,\kappa,\lambda) \in A^k\right\}, \\
D^k_{\tau m} &= \left\{(\tau,m) : \text{there exists } \eta \text{ such that } (\eta,\tau,m) \in D^k\right\}.
\end{aligned}$$

By Lemma 5, there is a one-to-one mapping between $\theta \in A^k$ and $d \in D^k$.

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 $\delta(\tau, m), \kappa(\tau, m), \lambda(\eta, \tau, m)$ denote the value of δ, κ, λ for given (η, τ, m), and let $\eta(\delta, \kappa, \lambda), \tau(\delta, \kappa), m(\delta, \kappa)$ denote the value of (η, τ, m) for given (δ, κ).

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

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

so that

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

Finally, let

$$LM_n^d(\eta, \tau, m) = 2\mathcal{G}_n^d(\tau, m)\sqrt{n}\eta - V(\tau, m)n\eta^2,$$
$$LR_n^d(\eta, \tau, m) = LR_n(\delta(\tau, m), \kappa(\tau, m), \lambda(\eta, \tau, m)),$$

There will be four steps in the proof:

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

$$LR_n^d(\eta_n, \tau_n, m_n) - LM_n^d(\eta_n, \tau_n, m_n) = o_p(n\eta_n^2).$$

2. $\left\{\mathcal{G}_n^d(\tau, m) : (\tau, m) \in D^1_{\tau m}\right\}$ is Donsker.

3. We prove that

$$\sup_{d \in D^1} LR_n^d(d) = \sup_{d \in D^1} LM_n^d(d) + o_p(1) = \sup_{(\tau, m) \in D^1_{\tau m}} \frac{\left(\min\left\{\mathcal{G}_n^d(\tau, m), 0\right\}\right)^2}{V(\delta, \kappa)} + o_p(1).$$

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

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

Step 1

Lemma 2 *Let $R_n^d(\eta, \tau, m) = LR_n^d(\eta, \tau, m) - LM_n^d(\eta, \tau, m)$. For all sequences of $(\eta_n, \tau_n, m_n) \in D^1$ and $\eta_n \xrightarrow{p} 0$, we have that*

$$R_n^d(\eta_n, \tau_n, m_n) = o_p\left(n\eta_n^2\right).$$

Proof. Let $\delta_n = \delta(\tau_n, m_n)$, $\kappa_n = \kappa(\tau_n, m_n)$, $\lambda_n = \lambda(\eta_n, \tau_n, m_n)$. 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 $(1-\lambda_n) \le \sqrt{\eta_n}$ or

$$\max\left\{\left|\frac{1}{36}\delta_n^4 - \frac{1}{8}\kappa_n^2\right|, \left|\frac{1}{2}\delta_n\kappa_n\right|\right\} \le \sqrt{\eta_n}. \tag{63}$$

Under (63), we have

$$2\eta \ge \left(\frac{1}{36}\delta_n^4 - \frac{1}{8}\kappa_n^2\right)^2 + \frac{1}{4}\delta_n^2\kappa_n^2 = \left(\frac{1}{36}\delta_n^4\right)^2 + \left(\frac{1}{8}\kappa_n^2\right)^2 + \frac{1}{4}\delta_n^2\kappa_n^2\left(1 - \frac{1}{36}\delta_n^2\right). \tag{64}$$

It is then easy to verify that given (63), $1 - \frac{1}{36}\delta_n^2 \ge 0$ with probability approaching 1. Therefore, (64) implies that

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

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

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

Second, a third-order Taylor expansion gives

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

The first term is

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

In turn, the second term will be

$$\frac{1}{2}\left\{\frac{1}{n}\frac{\partial^2 L_n(\delta_n,\kappa_n,1)}{\lambda^2}\right\}n(\lambda_n-1)^2 = \frac{1}{2}\left\{E\left[\frac{\partial^2 l(\delta_n,\kappa_n,1)}{\partial\lambda^2}\right]+O_p\left(\frac{\tau_n}{\sqrt{n}}\right)\right\}n(\lambda_n-1)^2 \tag{65}$$

$$= \frac{1}{2}E\left[\frac{\partial^2 l(\delta_n,\kappa_n,1)}{\partial\lambda^2}\right]n(\lambda_n-1)^2 + O_p[\sqrt{n}\tau_n(\lambda_n-1)^2]$$

$$= \frac{1}{2}E\left[\tau_n^{-2}\frac{\partial^2 l(\delta_n,\kappa_n,1)}{\partial\lambda^2}\right]n\tau_n^2(\lambda_n-1)^2 + O_p[\sqrt{n}\tau_n(\lambda_n-1)^2] \tag{66}$$

$$= -\frac{1}{2}V^d(\tau_n,m_n)n\tau_n^2(\lambda_n-1)^2 + o_p[\sqrt{n}\tau_n(\lambda_n-1)], \tag{67}$$

where (65) follows from Lemma 10(10.1); and (66) to (67) follow from the information matrix equality.

Let us now turn to the third term. By Lemmas 10(10.2) and 10(10.5), we have

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

$$= O(\tau_n)+O_p\left(\frac{1}{\sqrt{n}}\right),$$

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

$$\mathcal{K}[(\tau,m),(\tau',m')] = \frac{1}{\tau\tau'}cov\left\{\frac{\partial l[\delta(\tau,m),\kappa(\tau,m),1]}{\partial\lambda},\frac{\partial l[\delta(\tau',m'),\kappa(\tau',m'),1]}{\partial\lambda}\right\}. \tag{68}$$

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,m) 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 $(\tau,m)\subset\left[0,\bar{\delta}^4+\bar{\kappa}^2+\bar{\delta}\bar{\kappa}\right]\times[0,1]$.

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

Moreover,

$$E\left[\sup_{(\tau,m)\in D^1_{\tau m}}\left|\frac{1}{\tau}\frac{\partial l(\delta(\tau,m),\kappa(\tau,m),1)}{\partial\lambda}\right|\right]\leq E\left[\sup_{|\delta|\leq\bar{\delta}^2,|\kappa|\leq\bar{\kappa}^2,\delta^2+\kappa^2>0}\left|\frac{1}{\tau(\delta,\kappa)}\frac{\partial l(\delta,\kappa,1)}{\partial\lambda}\right|\right]<\infty. \tag{69}$$

To prove (69), consider the fifth-order Taylor expansion of $\frac{\partial l(\delta,\kappa,1)}{\partial\lambda}$ around $(\delta,\kappa)=(0,0)$ given by

$$\begin{aligned}\frac{\partial l(\delta,\kappa,1)}{\partial\lambda}=&\sum_{k=1}^{4}\sum_{i+j=k}\frac{1}{i!j!}\frac{\partial^{1+k}l(0,0,1)}{\partial\lambda\partial\delta^i\partial\kappa^j}\delta^i\kappa^j+\sum_{i+j=5}\frac{1}{i!j!}\frac{\partial^6 l\left(\tilde{\delta},\tilde{\kappa},1\right)}{\partial\lambda\partial\delta^i\partial\kappa^j}\delta^i\kappa^j\\ =&h_4\left(\frac{1}{36}\delta^4-\frac{1}{8}\kappa^2\right)+h_3\frac{1}{2}\delta\kappa+\sum_{4\geq i+j\geq 3,i\geq1,j\geq1}\frac{1}{i!j!}\frac{\partial^6 l(0,0,1)}{\partial\lambda\partial\delta^i\partial\kappa^j}\delta^i\kappa^j\\ &+\sum_{i+j=5,i\geq1,j\geq1}\frac{1}{i!j!}\frac{\partial^6 l\left(\tilde{\delta},\tilde{\kappa},1\right)}{\partial\lambda\partial\delta^i\partial\kappa^j}\delta^i\kappa^j+\frac{\partial^6 l\left(\tilde{\delta},\tilde{\kappa},1\right)}{\partial\lambda\partial\delta^5}\delta^5\\ &+\left(\frac{\partial^4 l(0,0,1)}{\partial\lambda\partial\kappa^3}+\frac{\partial^5 l(0,0,1)}{\partial\lambda\partial\kappa^4}\kappa+\frac{\partial^6 l\left(\tilde{\delta},\tilde{\kappa},1\right)}{\partial\lambda\partial\kappa^5}\kappa^2\right)\kappa^3.\end{aligned} \tag{70}$$

Consequently

$$\begin{aligned}\left|\frac{1}{\tau(\delta,\kappa)}\frac{\partial L(\delta,\kappa,1)}{\partial\lambda}\right|\leq&|h_4|+|h_3|+\sum_{4\geq i+j\geq3,i\geq1,j\geq1}\left|\frac{\partial^6 l(0,0,1)}{\partial\lambda\partial\delta^i\partial\kappa^j}\right|\frac{2}{i!j!}\bar{\delta}^{i-1}\bar{\kappa}^{j-1}\\ +&\sum_{i+j=5,i\geq1,j\geq1}\sup_{|\tilde{\delta}|\leq\bar{\delta},|\tilde{\kappa}|\leq\bar{\kappa}}\left|\frac{1}{i!j!}\frac{\partial^6 l\left(\tilde{\delta},\tilde{\kappa},1\right)}{\partial\lambda\partial\delta^i\partial\kappa^j}\right|\bar{\delta}^{i-1}\bar{\kappa}^{j-1}+\sup_{|\tilde{\delta}|\leq\bar{\delta},|\tilde{\kappa}|\leq\bar{\kappa}}\left|\frac{\partial^6 l\left(\tilde{\delta},\tilde{\kappa},1\right)}{\partial\lambda\partial\delta^5}\right|\left|\frac{\delta^5}{\tau(\delta,\kappa)}\right|\\ +&\left(\left|\frac{\partial^4 l(0,0,1)}{\partial\lambda\partial\kappa^3}\right|+\left|\frac{\partial^5 l(0,0,1)}{\partial\lambda\partial\kappa^4}\right||\bar{\kappa}|+\sup_{|\tilde{\delta}|\leq\bar{\delta},|\tilde{\kappa}|\leq\bar{\kappa}}\left|\frac{\partial^6 l\left(\tilde{\delta},\tilde{\kappa},1\right)}{\partial\lambda\partial\kappa^5}\right|\bar{\kappa}^2\right)\left|\frac{\kappa^3}{\tau(\delta,\kappa)}\right|.\end{aligned} \tag{71}$$

It is then easy to check that

$$E\left[|h_4|+|h_3|+\sum_{4\geq i+j\geq3,i\geq1,j\geq1}\frac{1}{i!j!}\left|\frac{\partial^6 l(0,0,1)}{\partial\lambda\partial\delta^i\partial\kappa^j}\right|+\left|\frac{\partial^4 l(0,0,1)}{\partial\lambda\partial\kappa^3}\right|+\left|\frac{\partial^5 l(0,0,1)}{\partial\lambda\partial\kappa^4}\right|\right]<\infty, \tag{72}$$

and

$$E\left[\sum_{i+j=5,i\geq0,j\geq0}\sup_{|\tilde{\delta}|\leq\bar{\delta},|\tilde{\kappa}|\leq\bar{\kappa}}\left|\frac{\partial^6 l\left(\tilde{\delta},\tilde{\kappa},1\right)}{\partial\lambda\partial\delta^i\partial\kappa^j}\right|\right]<\infty. \tag{73}$$

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

$$\frac{\kappa^2}{\tau(\delta,\kappa)}=\frac{1}{\max\left\{\left|\frac{1}{36}\frac{\delta^2}{\kappa^2}\delta^2-\frac{1}{8}\right|,\left|\frac{1}{2}\frac{\delta}{\kappa}\right|\right\}}\leq\begin{cases}\frac{2}{\left|\frac{\delta}{\kappa}\right|}\leq 2\bar{\delta} & \text{if } \frac{\delta^2}{\kappa^2}\geq\frac{1}{\bar{\delta}^2}\\ \frac{1}{\left|\frac{1}{36}\frac{\delta^2}{\kappa^2}\delta^2-\frac{1}{8}\right|}\leq\frac{1}{\left|\frac{1}{36}-\frac{1}{8}\right|}=\frac{72}{7} & \text{if } \frac{\delta^2}{\kappa^2}\leq\frac{1}{\bar{\delta}^2}.\end{cases} \tag{74}$$

Finally,

$$\left|\frac{\delta^5}{\tau}\right| \le \delta\left(\frac{\left|\delta^4-\frac{36}{8}\kappa^2\right|}{\max\left\{\left|\frac{1}{36}\delta^4-\frac{1}{8}\kappa^2\right|,\left|\frac{1}{2}\delta\kappa\right|\right\}}+\frac{\frac{36}{8}\kappa^2}{\max\left\{\left|\frac{1}{36}\delta^4-\frac{1}{8}\kappa^2\right|,\left|\frac{1}{2}\delta\kappa\right|\right\}}\right) < \bar{\delta}\left(1+2\bar{\delta}+\frac{72}{7}\right) \quad (75)$$

In sum, (69) follows from (71) - (75). But given (69), 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 (68).

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

$$\limsup_{n\to\infty} P\left[\sup_{\|(\tau_1,m_1)-(\tau_2,m_2)\|\le c,\ (\tau_1,m_1),(\tau_2,m_2)\in D^1_{\tau m}}\left|\mathcal{G}_n^d(\tau_1,m_1)-\mathcal{G}_n^d(\tau_2,m_2)\right| > \varepsilon\right] < \varepsilon.$$

In the rest of this section, we keep the restriction $(\tau_1, m_1), (\tau_2, m_2) \in D^1_{\tau m}$ implicit to simplify notation. First note that for $0 < c \le c_1$,

$$\begin{aligned}
&\sup_{\|(\tau_1,m_1)-(\tau_2,m_2)\|\le c}\left|\mathcal{G}_n^d(\tau_1,m_1)-\mathcal{G}_n^d(\tau_2,m_2)\right| \qquad (76)\\
&\le \max\Bigg\{\sup_{\|(\tau_1,m_1)-(\tau_2,m_2)\|\le c,|\tau_1|,|\tau_2|\le 2c_1}\left|\mathcal{G}_n^d(\tau_1,m_1)-\mathcal{G}_n^d(\tau_2,m_2)\right|,\\
&\qquad \sup_{\|(\tau_1,m_1)-(\tau_2,m_2)\|\le c,|\tau_1|,|\tau_2|\ge c_1}\left|\mathcal{G}_n^d(\tau_1,m_1)-\mathcal{G}_n^d(\tau_2,m_2)\right|\Bigg\},
\end{aligned}$$

whence

$$\begin{aligned}
&P\left[\sup_{\|(\tau_1,m_1)-(\tau_2,m_2)\|\le c}\left|\mathcal{G}_n^d(\tau_1,m_1)-\mathcal{G}_n^d(\tau_2,m_2)\right| > \varepsilon\right]\\
\le &P\Bigg[\left\{\sup_{\|(\tau_1,m_1)-(\tau_2,m_2)\|\le c,|\tau_1|,|\tau_2|\le 2c_1}\left|\mathcal{G}_n^d(\tau_1,m_1)-\mathcal{G}_n^d(\tau_2,m_2)\right| > \varepsilon\right\}\\
&\quad \cup\left\{\sup_{\|(\tau_1,m_1)-(\tau_2,m_2)\|\le c,|\tau_1|,|\tau_2|\ge c_1}\left|\mathcal{G}_n^d(\tau_1,m_1)-\mathcal{G}_n^d(\tau_2,m_2)\right| > \varepsilon\right\}\Bigg]\\
\le &P\left[\sup_{\|(\tau_1,m_1)-(\tau_2,m_2)\|\le c,|\tau_1|,|\tau_2|\le 2c_1}\left|\mathcal{G}_n^d(\tau_1,m_1)-\mathcal{G}_n^d(\tau_2,m_2)\right| > \varepsilon\right] \qquad (77)\\
&\quad + P\left[\sup_{\|(\tau_1,m_1)-(\tau_2,m_2)\|\le c,|\tau_1|,|\tau_2|\ge c_1}\left|\mathcal{G}_n^d(\tau_1,m_1)-\mathcal{G}_n^d(\tau_2,m_2)\right| > \varepsilon\right]. \qquad (78)
\end{aligned}$$

For the first term in (77), we show that for all $\varepsilon > 0$, there exist $c_1 \ge c_2 > 0$ such that

$$P\left[\sup_{\|(\tau_1,m_1)-(\tau_2,m_2)\|\le c_2,|\tau_1|,|\tau_2|\le 2c_1}\left|\mathcal{G}_n^d(\tau_1,m_1)-\mathcal{G}_n^d(\tau_2,m_2)\right| > \varepsilon\right] \le \frac{\varepsilon}{2}.$$

Given (70), we will have that

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

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

$$\frac{1}{21}\left[\mathcal{G}_n^d(\tau_1,m_1)-\mathcal{G}_n^d(\tau_2,m_2)\right]^2$$
$$\leq \left(\frac{H_4}{\sqrt{n}}\right)^2\left\{\tau_1^{-1}\left(\frac{1}{36}\delta_1^4-\frac{1}{8}\kappa_1^2\right)-\tau_2^{-1}\left(\frac{1}{36}\delta_2^4-\frac{1}{8}\kappa_2^2\right)\right\}^2 \tag{79}$$
$$+\left(\frac{H_3}{\sqrt{n}}\right)^2\left\{\frac{1}{2}\tau_1^{-1}\delta_1\kappa_1-\frac{1}{2}\tau_2^{-1}\delta_2\kappa_2\right\}^2 \tag{80}$$
$$+\sum_{4\geq i+j\geq 3, j\geq 1}\left(\frac{1}{i!j!}\frac{1}{\sqrt{n}}\frac{\partial^6 L_n(0,0,1)}{\partial\lambda\partial\delta^i\partial\kappa^j}\right)^2\left\{\tau_1^{-1}\delta_1^i\kappa_1^j-\tau_2^{-1}\delta_2^i\kappa_2^j\right\}^2 \tag{81}$$
$$+\sum_{i+j=5}\sup_{|\delta|\leq\bar{\delta},|\kappa|\leq\bar{\kappa}}\left(\frac{1}{i!j!}\frac{1}{\sqrt{n}}\frac{\partial^6 L_n(\delta,\kappa,1)}{\partial\lambda\partial\delta^i\partial\kappa^j}\right)^2\left\{\tau_1^{-2}\delta_1^{2i}\kappa_1^{2j}+\tau_2^{-2}\delta_2^{2i}\kappa_2^{2j}\right\}, \tag{82}$$

where $\delta_1 = \delta(\tau_1, m_1)$, $\kappa_1 = \kappa(\tau_1, m_1)$, δ_2 and κ_2 are defined in the same way. First, we can easily check that

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

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

by the *iid* assumption and the zero expectation of these terms. Second, for the terms (79)-(82), we can show that the non-random coefficients in {} converge to zero as $c, c_1 \to 0$, using arguments in (74), (75) and Lemma 11. To be more specific, for $(\tau, m) \in B^1$, we have

$$\tau_1^{-1}\left(\frac{1}{36}\delta_1^4-\frac{1}{8}\kappa_1^2\right)-\tau_2^{-1}\left(\frac{1}{36}\delta_2^4-\frac{1}{8}\kappa_2^2\right) = 1-1=0$$
$$\frac{1}{2}\tau_1^{-1}\delta_1\kappa_1-\frac{1}{2}\tau_2^{-1}\delta_2\kappa_2 = \frac{1}{2}(m_1-m_2)$$
$$\tau_1^{-1}\delta_1^i\kappa_1^j-\tau_2^{-1}\delta_2^i\kappa_2^j = \begin{cases} = m_1\delta_1^{i-1}\kappa_1^{j-1}-m_2\delta_2^{i-1}\kappa_2^{j-1} & \text{if } i\geq 1 \\ = \tau_1^{-1}\kappa_1^j-\tau_2^{-1}\kappa_2^j \leq \sup\left|\frac{\kappa^2}{\tau}\right|(\kappa_1+\kappa_2) & \text{if } i=0\end{cases},$$

and the same applies to $\tau_1^{-2}\delta_1^{2i}\kappa_1^{2j}$. Together with Lemma 10(10.3), 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

$$E\left[\sup_{\|(\tau_1,m_1)-(\tau_2,m_2)\|\leq c_2,\tau_1,\tau_2\leq 2c_1}\left(\mathcal{G}_n^d(\tau_1,m_1)-\mathcal{G}_n^d(\tau_2,m_2)\right)^2\right]\leq\frac{\varepsilon^3}{2}. \tag{83}$$

Then Chebychev's inequality implies that

$$\begin{aligned}&P\left[\sup_{\|(\tau_1,m_1)-(\tau_2,m_2)\|\leq c,|\tau_1|,|\tau_2|\leq 2c_1}\left|\mathcal{G}_n^d(\tau_1,m_1)-\mathcal{G}_n^d(\tau_2,m_2)\right|>\varepsilon\right]\\&\quad\leq\frac{1}{\varepsilon^2}E\left[\sup_{\|(\tau_1,m_1)-(\tau_2,m_2)\|\leq c,|\tau_1|,|\tau_2|\leq 2c_1}\left(\mathcal{G}_n^d(\tau_1,m_1)-\mathcal{G}_n^d(\tau_2,m_2)\right)^2\right]\leq\varepsilon.\end{aligned}$$

Next, consider (78). Given c_1, we need to find c such that $c_1 \geq c > 0$ and

$$P\left[\sup_{\|(\tau_1,m_1)-(\tau_2,m_2)\|\leq c,|\tau_1|,|\tau_2|\geq c_1}\left|\mathcal{G}_n^d(\tau_1,m_1)-\mathcal{G}_n^d(\tau_2,m_2)\right|>\varepsilon\right]\leq\frac{\varepsilon}{2}. \tag{84}$$

First, we change (τ,m) to (δ,κ) for simplicity. For $(\tau,m)\in D^1$, it holds that

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

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

$$\begin{aligned}&\left\{(\tau_1,m_1,\tau_2,m_2)\in B_{\tau m}^1\times B_{\tau m}^1:\|(\tau_1,m_1)-(\tau_2,m_2)\|\leq c,\tau_1,\tau_2\geq c_1\right\}\\&\quad\subset\left\{(\tau_1,m_1,\tau_2,m_2)\in B_{\tau m}^1\times B_{\tau m}^1:\|(\delta_1,\kappa_1)-(\delta_2,\kappa_2)\|\leq c_B,\delta_1,\delta_2\geq\sqrt{6}c_1^{\frac{1}{4}}\right\}\end{aligned} \tag{85}$$

because $\left\{(\tau,m)\in D_{\tau m}^1:\tau\geq c_1\right\}$ is a compact set, and $\tau(\delta,\kappa)$ and $m(\delta,\kappa)$ are continuous on this set. Therefore, it suffices to find c_B such that $\left\{\mathcal{G}_n(\delta,\kappa):|\delta|\geq\sqrt{6}c_1^{1/4},(\delta,\kappa)\in A_{\delta\kappa}^1\right\}$ is stochastically equicontinuous. 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

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

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

Note that

$$
\begin{aligned}
\frac{1}{\tau_1}\frac{\partial l}{\partial \lambda}(\tau_1, m_1) - \frac{1}{\tau_2}\frac{\partial l}{\partial \lambda}(\tau_2, m_2) = {} & y^2\left[D_1(\tau_1,\delta_1,\kappa_1) - D_1(\tau_2,\delta_2,\kappa_2)\right] \\
& + y\left[D_2(\tau_1,\delta_1,\kappa_1) - D_2(\tau_2,\delta_2,\kappa_2)\right] \\
& + \left[D_3(\tau_1,\delta_1,\kappa_1) - D_3(\tau_2,\delta_2,\kappa_2)\right] \\
& - \frac{1}{\tau_1}\exp\left[-\frac{e^{\frac{\delta_1^2}{3}-\kappa_1}}{2}(\delta_1+y)^2 + \frac{1}{2}y^2 + \frac{1}{6}\delta_1^2 - \frac{1}{2}\kappa_1\right] \\
& + \frac{1}{\tau_2}\exp\left[-\frac{e^{\frac{\delta_2^2}{3}-\kappa_2}}{2}(\delta_2+y)^2 + \frac{1}{2}y^2 + \frac{1}{6}\delta_2^2 - \frac{1}{2}\kappa_2\right],
\end{aligned} \tag{86}
$$

where

$$
D_1(\tau,\delta,\kappa) = \frac{1}{2}\tau^{-1}e^{\kappa-\frac{\delta^2}{3}} + \frac{1}{2}\frac{\delta^2}{\tau}, \qquad D_2(\tau,\delta,\kappa) = -\frac{\delta}{\tau},
$$

$$
D_3(\tau,\delta,\kappa) = -\frac{1}{2}\tau^{-1}\left(e^{\kappa-\frac{\delta^2}{3}} - \delta^2\right)
$$

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 (86), the mean value theorem implies that

$$
\begin{aligned}
& -\frac{1}{\tau_1}\exp\left[-\frac{e^{\frac{\delta_1^2}{3}-\kappa_1}}{2}(\delta_1+y)^2 + \frac{1}{2}y^2 + \frac{1}{6}\delta_1^2 - \frac{1}{2}\kappa_1\right] \\
& + \frac{1}{\tau_2}\exp\left[-\frac{e^{\frac{\delta_2^2}{3}-\kappa_2}}{2}(\delta_2+y)^2 + \frac{1}{2}y^2 + \frac{1}{6}\delta_2^2 - \frac{1}{2}\kappa_2\right] \\
= {} & \exp\left[-\frac{e^{\frac{\tilde{\delta}^2}{3}-\tilde{\kappa}}}{2}(\tilde{\delta}+y)^2 + \frac{1}{2}y^2 + \frac{1}{6}\tilde{\delta}^2 - \frac{1}{2}\tilde{\kappa}\right]\left\{\frac{1}{\tilde{\tau}^2}(\tau_1-\tau_2)\right. \\
& + \frac{1}{3\tilde{\tau}}\left[e^{\frac{\tilde{\delta}^2}{3}-\tilde{\kappa}}\left(\tilde{\delta}^3 + 3\tilde{\delta} + \tilde{\delta}y^2 + 2\tilde{\delta}^2 y + 3y\right) - \tilde{\delta}\right](\delta_1-\delta_2) \\
& \left. + \frac{1}{2\tilde{\tau}}\left[1 - e^{\frac{\tilde{\delta}^2}{3}-\tilde{\kappa}}(\tilde{\delta}+y)^2\right](\kappa_1-\kappa_2)\right\}.
\end{aligned} \tag{87}
$$

In addition,

$$
\begin{aligned}
|\tau_1-\tau_2| &= \left|\frac{1}{36}\delta_1^4 - \frac{1}{8}\kappa_1^2 - \frac{1}{36}\delta_2^4 + \frac{1}{8}\kappa_2^2\right| \\
&= \left|\frac{1}{36}\left(\delta_1^2+\delta_2^2\right)(\delta_1+\delta_2)(\delta_1-\delta_2) - \frac{1}{8}(\kappa_1+\kappa_2)(\kappa_1-\kappa_2)\right| \\
&\le \frac{1}{9}\bar{\delta}^3\left|\delta_1-\delta_2\right| + \frac{\bar{\kappa}}{4}\left|\kappa_1-\kappa_2\right|.
\end{aligned} \tag{88}
$$

Moreover

$$\exp\left(-\frac{e^{\frac{\delta^2}{3}-\kappa}}{2}(\delta+y)^2+\frac{1}{2}y^2+\frac{1}{6}\delta^2-\frac{1}{2}\kappa\right)\le g^*(y), \tag{89}$$

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 (86), (87), (88) and (89), we will have

$$\frac{1}{\tau_1}\frac{\partial l}{\partial\lambda}(\tau_1,m_1)-\frac{1}{\tau_2}\frac{\partial l}{\partial\lambda}(\tau_2,m_2)\le(g^*(y)+1)\left\{c_1+c_2|y|+c_3y^2\right\}(|\delta_1-\delta_2|+|\kappa_1-\kappa_2|).$$

But since

$$E\left[(g^*(y)+1)\left\{c_1+c_2|y|+c_3y^2\right\}\right]<\infty,$$

f will be Lipschitz with $M(y)=(g^*(y)+1)\left(c_1+c_2|y|+c_3y^2\right)$ for some constants c_1, c_2 and c_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:

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

is also satisfied because y_i is a standard normal r.v. 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 (84) holds.

In sum, we have

$$\begin{aligned}
&\limsup_{n\to\infty}P\left[\sup_{\|(\tau_1,m_1)-(\tau_2,m_2)\|\le c}\left|\mathcal{G}_n^d(\tau_1,m_1)-\mathcal{G}_n^d(\tau_2,m_2)\right|>\varepsilon\right]\\
\le&\limsup_{n\to\infty}P\left[\sup_{\|(\tau_1,m_1)-(\tau_2,m_2)\|\le c_2,\ |\tau_1|,|\tau_2|\le 2c_1}\left|\mathcal{G}_n^d(\tau_1,m_1)-\mathcal{G}_n^d(\tau_2,m_2)\right|>\varepsilon\right]\\
&\quad+\limsup_{n\to\infty}P\left[\sup_{\|(\tau_1,m_1)-(\tau_2,m_2)\|\le c,\ |\tau_1|,|\tau_2|\ge c_1}\left|\mathcal{G}_n^d(\tau_1,m_1)-\mathcal{G}_n^d(\tau_2,m_2)\right|>\varepsilon\right]\\
\le&\frac{\varepsilon}{2}+\frac{\varepsilon}{2}\le\varepsilon,
\end{aligned}$$

as desired. □

Step 3

Lemma 4 $\sup_{d\in D^1} LR_n^d(d) = \sup_{d\in D^1} LM_n^d(d) + o_p(1) = \sup_{(\tau,m)\in D^1_{\tau m}} \frac{\left[\mathcal{G}_n^d(\tau,m)\right]_-^2}{V^d(\tau,m)} + o_p(1).$

Proof. Since

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

it suffices to show that

$$\sup_{\eta:(\eta,\tau,m)\in D^1} LR_n^d(\eta,\tau,m) = \sup_{\eta:(\eta,\tau,m)\in D^1} LM_n^d(\eta,\tau,m) + o_p(1). \tag{90}$$

Expression (90) follows from Andrews (2001). To see this, we need to check his assumptions. Let

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

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

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

To verify Assumption 1, namely $\hat{\eta}_{\tau m} = o_{p,\tau m}(1)$, let $l_0^d(d) = E\left[l^d(1,\tau,m)\right]$. Invoking Lemma 8, we have

$$\sup_{b\in B^1}\left|\frac{1}{n}L_n^d(d) - l_0^d(0,\tau,m)\right| \leq \sup_{\theta\in\Theta}\left|\frac{1}{n}L_n(\theta) - l_0(\theta)\right| \xrightarrow{p} 0 \tag{91}$$

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

$$l_0^d(d) > \sup_{\eta>\epsilon, d\in \mathrm{cl}(B^1)} l_0^d(d) \tag{92}$$

(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 $\hat{\eta}_{\tau m} = o_{p,\tau m}(1)$.

Assumption 2* holds with $B_T = \sqrt{n}$, see Lemma 2. Assumption 3* holds by Lemma 3. Assumption 4 is implied by Assumptions 1, 2* and 3. 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 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

(90) holds. $\square$

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 $\mathcal{V}$, whereas L_n is the log-likelihood indexed by θ. First, by the results in Step 3, we have

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

Noticing also that

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

we will have that

$$\begin{aligned}\sup_{\theta\in P_b} LR_n(\theta) &= \max_{k\leq 16}\sup_{b\in B^k} LR_n^d(b)=\max_{k\leq 16}\sup_{(\delta,\kappa)\in A_{\tau m}^k}\frac{\left\lfloor \mathcal{G}_n(\delta,\kappa)\right\rfloor_-^2}{V(\delta,\kappa)}+o_p(1)\\ &= \sup_{(\delta,\kappa):(\delta,\kappa,1)\in P_b}\frac{\left\lfloor \mathcal{G}_n(\delta,\kappa)\right\rfloor_-^2}{V(\delta,\kappa)}+o_p(1).\end{aligned}$$

Therefore,

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

Proof of Proposition 5

To show (8), note that for $k_1\in\mathbb{R}$,

$$\lim_{\varepsilon\to 0}\left\lfloor\frac{\mathcal{G}_n(\varepsilon,k_1\varepsilon)}{\sqrt{V(\varepsilon,k_1\varepsilon)}}\right\rfloor_-^2=\frac{1}{n}\frac{\left\lfloor 4k_1H_{3,n}-k_1^2H_{4,n}\right\rfloor_-^2}{16k_1^2V_3+k_1^4V_4}. \tag{93}$$

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

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

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

$$
\lim_{\varepsilon\to 0}\left\lfloor \frac{\mathcal{G}_n(\varepsilon, \frac{k_2}{18}\varepsilon^3)}{\sqrt{V(\varepsilon, \frac{k_2}{18}\varepsilon^3)}}\right\rfloor_-^2 = \frac{1}{n}\frac{\lfloor H_{4,n} + k_2 H_{3,n}\rfloor_-^2}{V_4 + k_2^2 V_3}. \tag{94}
$$

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

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

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

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

In turn, to show (9), note that

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

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,

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

as desired. □

Proof of Proposition 6

(a) By LeCam'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

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

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

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

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 (a) holds.

(b) Using the results from (a), 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

$$\mathcal{N}\left[\begin{pmatrix}0\\0\\0\\-\frac{1}{2}var[G(\beta)]\rho^2\end{pmatrix},\begin{pmatrix}V_3&.&.&.\\0&V_4&.&.\\c_3&c_4&var[G(\beta)]&.\\-c_3\rho&-c_4\rho&-var[G(\beta)]\rho&var[G(\beta)]\rho^2\end{pmatrix}\right].\tag{95}$$

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

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

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

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

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 (95), it follows from LeCam'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 H_{1n} to a normal distribution with mean $E(T) + cov[T, \ln(U)]$ and the same variance $V(T)$ as under H_0, which proves result (b).

Part (c) then follows from the joint distribution of $\left(\frac{H_{3,n}}{\sqrt{n}}, \frac{H_{4,n}}{\sqrt{n}}\right)$ under H_{1n} derived in (b).

Finally, the limiting distribution of $\widetilde{LM}_{b,1}$ test in part (d) follows from the continuous mapping theorem. □

Proof of Proposition 7

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 $\ln\frac{dP_{\theta_n}}{dP_{\theta_0}}$ under P_{θ_0}. It is easy to see that

$$\begin{pmatrix} \frac{H_{3,n}}{\sqrt{n}} \\ \frac{H_{4,n}}{\sqrt{n}} \\ \frac{1}{\sqrt{n}}\frac{\partial L_n(\beta,1)}{\partial\lambda} \\ \ln\frac{dP_{\theta_n}}{dP_{\theta_0}} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ -\frac{V_3 w_1^2}{2}-\frac{V_4 w_2^2}{2} \end{pmatrix}, \begin{pmatrix} V_3 & . & . & . \\ 0 & V_4 & . & . \\ c_3 & c_4 & var[G(\beta)] & . \\ V_3 w_1 & V_4 w_2 & c_3 w_1 + c_4 w_2 & V_3 w_1^2 + V_4 w_2^2 \end{pmatrix}\right]$$

under P_{θ_0}. It then follows from Le Cam's third lemma (see van der Vaart (1998)) 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 P_{θ_n}. Therefore,

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

as desired. □

Proof of Proposition 8

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 given simply 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_i 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 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 $\theta = (\delta,\kappa,\lambda)$. Next, define

$$LR_n(\mu,\sigma^2,\delta,\kappa,\lambda) = 2\left[L_n(\mu,\sigma^2,\delta,\kappa,\lambda) - L_n(\mu_0,\sigma_0^2,0,0,\lambda)\right] \tag{96}$$

and

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

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

Let

$$\begin{aligned} LM_n^{a,\phi}(\phi) =& 2\left(\frac{1}{\sigma_0}\frac{H_{1,n}}{\sqrt{n}}\right)\sqrt{n}(\mu-\mu_0) + 2\left(\frac{1}{2\sigma_0^2}\frac{H_{2,n}}{\sqrt{n}}\right)\sqrt{n}(\sigma^2-\sigma_0^2) \\ & - \frac{1}{\sigma_0^2}n(\mu-\mu_0)^2 - \frac{1}{2\sigma_0^4}n\left(\sigma^2-\sigma_0^2\right)^2, \end{aligned} \tag{97}$$

where

$$\begin{aligned} H_{1,n} &= \sum_{i=1}^{n} h_{1i} = \sum_{i=1}^{n}\frac{y_i-\mu_0}{\sqrt{\sigma_0^2}}, \\ H_{2,n} &= \sum_{i=1}^{n} h_{2i} = \sum_{i=1}^{n}\frac{(y_i-\mu_0)^2-\sigma_0^2}{\sigma_0^2}, \end{aligned}$$

so that $LM_n^a(\theta;\phi_0)$ coincides with (10) 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 $\rho_{n,u}^{LR} = (\phi_{n,u}^{LR}, \delta_{n,u}^{LR}, \kappa_{n,u}^{LR}, \lambda_{n,u}^{LR}) \in \arg\max_{\phi\in\Phi\times P} LR_n(\rho)$, 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 replace y_i by $\frac{y_i-\mu_{n,r}}{\sigma_{n,r}}$.

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

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

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,

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

where

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

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

$$
\frac{\sqrt{n}}{n}\left.\frac{\partial^{k_1+k_2+k_3+k_4}L_n(\rho)}{\partial\mu^{k_1}\left(\partial\sigma^2\right)^{k_2}\partial\delta^{k_3}\partial\kappa^{k_4}}\right|_{\rho_0} = O_p(1). \tag{98}
$$

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

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

where the last equality follows from (98) 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

$$
n\left(\mu_n-\mu_0\right)^{k_1}\left(\sigma_n^2-\sigma_0^2\right)^{k_2}\delta_n^{k_3}\kappa_n^{k_4} = o_p[h^\phi(\phi_n)], \tag{99}
$$

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

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

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

$$n\mu_n^{k_1}\left(\sigma_n^2\right)^{k_2}(1-\lambda_n)\delta_n^{k_3}\kappa_n^{k_4} = \sqrt{n}\left(\mu_n-\mu_0\right)^{k_1}\left(\sigma_n^2-\sigma_0^2\right)^{k_2}\sqrt{n}(1-\lambda_n)\delta_n^{k_3}\kappa_n^{k_4} = o_p[h^\theta(\theta_n)]. \quad (100)$$

Secondly, we have to show that $T_{2n} = LM_n^\phi(\phi_n;\phi_0) + o_p[h^\phi(\phi_n)]$. 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

$$\begin{aligned}2\sum_{k_1+k_2=2} L_n^{[k_1,k_2,0,0]}\left(\mu_n-\mu_0\right)^{k_1}\left(\sigma_n^2-\sigma_0^2\right)^{k_2} &= 2\sum_{k_1+k_2=2}\frac{1}{n}L_n^{[k_1,k_2,0,0]}n\left(\mu_n-\mu_0\right)^{k_1}\left(\sigma_n^2-\sigma_0^2\right)^{k_2}\\ &= -\frac{V_1}{\sigma_0^2}n\left(\mu_n-\mu_0\right)^2 - \frac{V_2}{4\sigma_0^2}n\left(\sigma_n^2-\sigma_0^2\right)^2 + o_p[h^\phi(\phi_n)].\end{aligned}$$

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

Third, we have to show that $T_{1n} = LM_n(\theta_n) + o_p[h^\theta(\theta_n)]$. But since this is the same as we did in 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 $\max_{\phi\in\Phi} LM^\phi(\phi)$ has a closed-form solution with probability approaching one. The asymptotic properties of θ_n^{LM}

Step 3: Following the proof of Proposition 1, we can first show that $\rho_n^{LR} \xrightarrow{p} 0$. Next, we can also show that $h_n^\theta(\theta_n^{LR}) = O_p(1)$ and $h_n^\phi(\phi_n^{LR}) = O_p(1)$ (similar to Lemma 3 in Amengual, Bei and Sentana (2020)).

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

Step 5: Simplify $LM_n^\theta(\theta_n^{LM})$ is 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 P_b In terms of Andrews (2001) notation, we have

$$\beta_1 = \eta, \pi = (\tau, m), \psi = (\mu, \sigma^2).$$

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

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

where $LR_n^d(\mu,\sigma^2,\eta,\tau,m)$ is defined in (96) and $LM_n^{\phi}(\phi)$ in (97). We need to show that for all sequences $(\mu_n,\sigma_n^2,\eta_n,\tau_n,m_n)$ with $(\mu_n-\mu_0,\sigma_n^2-\sigma_0^2,\eta_n)\xrightarrow{p}0$, it holds that

$$
R_n(\mu_n,\sigma_n^2,\eta_n,\tau_n,m_n) = o_p\left\{\max[n\eta_n^2, n(\mu_n-\mu_0), n(\sigma_n^2-\sigma_0^2)^2]\right\}. \tag{101}
$$

To see this, we can modify the proof of Proposition 1. Let $\rho_n = (\mu_n,\sigma_n^2,\delta_n,\kappa_n,\lambda_n)$ with $\delta_n=\delta(\tau_n,m_n)$, $\kappa_n=\kappa(\tau_n,m_n)$ and $\lambda_n=\lambda(\eta_n,\tau_n,m_n)$. A third-order Taylor expansion gives

$$
\begin{aligned}
L^d(\mu_n,\sigma_n^2,\eta_n,\tau_n,m_n) - L^d(\mu_0,\sigma_0^2,0,\tau_n,m_n) &= L(\mu_n,\sigma_0^2+\sigma_n^2,\delta_n,\kappa_n,\lambda_n) - L(\sigma_0^2,\delta_n,\kappa_n,1)\\
&= T_{1n}(\rho_n;\phi_0) + T_{2n}(\rho_n;\phi_0)\\
&\quad + T_{3n}(\rho_n;\phi_0) + T_{4n}(\rho_n;\phi_0),
\end{aligned}
$$

where

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

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

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

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

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

$$
2T_{1n}(\rho_n;\phi_0) = 2\mathcal{G}_n(\tau_n,m_n)\sqrt{n}\eta_n - V(\tau_n,m_n)n\eta_n^2 + o_p(n\eta_n^2) \tag{102}
$$

using the same argument as in Proposition 1. Hence, it is straightforward to show that

$$2T_{2n}(\phi_n;\phi_0) = LM_n^{\phi}(\phi_n) + o_p\left[n\left(\sigma_n^2\right)^2 + n\mu_n^2\right] \tag{103}$$

We can also show that

$$\begin{aligned}
T_{3n}(\rho_n;\phi_0) =& \left\{\frac{1}{\sqrt{n}}\frac{\partial^2 L(\rho_{n0})}{\partial\lambda\partial\mu}\right\}\left[\sqrt{n}(\mu_n-\mu_0)\right](\lambda_n-1) \\
&+\left\{\frac{1}{\sqrt{n}}\frac{\partial^2 L(\rho_{n0})}{\partial\lambda\partial\sigma^2}\right\}\left[\sqrt{n}(\sigma_n^2-\sigma_0^2)\right](\lambda_n-1) \\
&+\frac{1}{2}\left\{\frac{1}{n}\tau_n^{-1}\frac{\partial^3 L(\tilde{\rho}_n)}{\partial\lambda^2\partial\mu}\right\}\left[n(\mu_n-\mu_0)\eta_n\right](\lambda_n-1) \\
&+\frac{1}{4}\left\{\frac{1}{n}\tau_n^{-1}\frac{\partial^3 L(\tilde{\rho}_n)}{\partial\lambda^2\partial\sigma^2}\right\}\left[n(\sigma_n^2-\sigma_0^2)\eta_n\right](\lambda_n-1) \\
=&o_p\left[n(\mu_n-\mu_0)^2 + n\left(\sigma_n^2-\sigma_0^2\right)^2 + n\eta_n^2\right],
\end{aligned} \tag{104}$$

where the first equality follows from $\eta_n = (\lambda_n - 1)\tau_n$ and the second one follows from Lemma 10 and $\lambda_n \xrightarrow{p} 1$. The last part is easy, as $n(\mu_n-\mu_0)^j(\sigma_n^2-\sigma_0^2)^k = O\left[n\mu_n^2 + n\left(\sigma_n^2\right)^2\right]$ and $\lambda_n \to 1$, so that

$$T_{4n} = o_p\left[n\left(\sigma_n^2\right)^2 + n\mu_n^2\right]. \tag{105}$$

Combining the results in (102), (103), (104) and (105), we finally prove (101).

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 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

$(k_2, k_3) \in \{(k, 0) : k \leq 4\} \cup \{(0, k) : k \leq 2\} \cup \{(1, 1)\}$,

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

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, the same arguments in Step 1 applies. Analogous arguments apply for the P_b part too, which completes the proof. □

Lemmas

Lemma 5 *For $k = 1, \ldots, 16$, let*

$$D^k = \Big\{(\eta, \tau, m) : \textit{there exists } \theta \in A^k \textit{ such that } (60)\text{-}(62) \textit{ holds}\Big\}.$$

Then, (i) for all $\theta \in A^k$, there exists a unique $d \in D^k$ such (60) - (62) holds; (ii) for all $d \in D^k$, there exists a unique $\theta \in A^k$ such that (60) - (62) 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 $\lambda = 1 - \frac{\eta}{\tau}$. With the restrictions of A^1, it holds that

$$\frac{1}{36}\delta^4 - \frac{1}{8}\kappa^2 = \tau, \ \text{ that is, } \frac{1}{2}\delta\kappa = m\tau. \tag{106}$$

Hence, we can easily write

$$\frac{2}{9}\delta^4 - \frac{4\tau^2 m^2}{\delta^2} = 8\tau. \tag{107}$$

Since the left hand side of (107) is strictly increasing in δ^2, we can get unique δ. Finally, we get κ from (106). □

Lemma 6

$$\left\{\left|w_{1n}^{LM}\right| \geq M, \left|\frac{n^{-\frac{1}{2}}H_{3,n}}{V_3}\right| < \frac{M_1}{\sqrt{3V_3}}, \left|\frac{n^{-\frac{1}{2}}H_{4,n}}{V_4}\right| < \frac{M_1}{\sqrt{3V_4}}\right\} = \emptyset$$

where
$$M = \frac{M_1}{\sqrt{V_3}}\left(1+\frac{1}{\sqrt{3}}\right).$$

Proof. It suffices to show that when

$$\left|\frac{n^{-\frac{1}{2}}H_{3,n}}{V_3}\right| < \frac{M_1}{\sqrt{3V_3}} \quad \text{and} \quad \left|\frac{n^{-\frac{1}{2}}H_{4,n}}{V_4}\right| < \frac{M_1}{\sqrt{3V_4}},$$

$$\sup_{\theta\in\Theta,|w_1|\geq M} LM_n(\theta) < LM_n(0,0,\lambda) = 0.$$

But

$$\begin{aligned}
&\sup_{\theta\in\Theta,|w_1|\geq M} LM_n(\theta)\\
=&\sup_{\theta\in\Theta,|w_1|\geq M}\left[-V_3\left(w_1-\frac{n^{-\frac{1}{2}}H_{3,n}}{V_3}\right)^2+\frac{H_{3,n}^2}{nV_3}-V_4\left(w_2-\frac{n^{-\frac{1}{2}}H_{4,n}}{V_4}\right)^2+\frac{H_{4,n}^2}{nV_4}\right]\\
\leq&\sup_{|w_1|\geq M}\left[-V_3\left(w_1-\frac{n^{-\frac{1}{2}}H_{3,n}}{V_3}\right)^2+\frac{H_{3,n}^2}{nV_3}+\frac{H_{4,n}^2}{nV_4}\right]\\
<&\sup_{|w_1|\geq M}\left[-V_3\left(w_1-\frac{n^{-\frac{1}{2}}H_{3,n}}{V_3}\right)^2\right]+\frac{2M_1^2}{3} \qquad (108)\\
<&-M_1^2+\frac{2M_1^2}{3}<0=LM_n(0,0,\lambda), \qquad (109)
\end{aligned}$$

which is a contradiction. Notice that from (108) to (109) we used the fact that when

$$|w_1| \geq M = \frac{M_1}{\sqrt{V_3}}\left(1+\frac{1}{\sqrt{3}}\right) \quad \text{and} \quad \left|\frac{n^{-\frac{1}{2}}H_{3,n}}{V_3}\right| < \frac{M_1}{\sqrt{3V_3}},$$

we have
$$\left(w_1-\frac{n^{-\frac{1}{2}}H_{3,n}}{V_3}\right)^2 > \frac{M_1^2}{V_3},$$

as desired. □

Lemma 7 *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 $1-\lambda_n+\lambda_n^2 \in [\frac{3}{4}, 1]$. 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 $\forall M_1$

$$\Pr(n^{\frac{1}{2}}(1-\lambda_n)\delta_n^4 > M_1) > \epsilon \text{ i.o.} \tag{110}$$

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 $M' > \max\{M_2, \frac{\bar{\delta}^2}{6}\}$ and let $M_1 = 6M' + 6M_2$. In view of (110), we have that

$$\Pr(n^{\frac{1}{2}}(1-\lambda_n)\delta_n^4 > 6M' + 6M_2) > \epsilon \text{ i.o.}$$

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 $\Pr(B_n) > 1 - \frac{\epsilon}{2}$ $\forall n$, we will also have

$$\Pr(A_n \cap B_n) \geq \Pr(A_n) + \Pr(B_n) - 1 > \frac{\epsilon}{2} \text{ i.o.}$$

Let us now consider the set $A_n \cap B_n$. We can prove that

$$\begin{aligned}
n(1-\lambda_n)^2\delta_n^2\kappa_n^2 &= \sqrt{n}(1-\lambda_n)\delta_n^2 \left\{\frac{2}{9}(1-\lambda_n+\lambda_n^2)\sqrt{n}(1-\lambda_n)\delta_n^4 \right. \\
&\quad \left. + \left[\sqrt{n}(1-\lambda_n)\kappa_n^2 - \frac{2}{9}(1-\lambda_n+\lambda_n^2)\sqrt{n}(1-\lambda_n)\delta_n^4\right]\right\} \\
&> \sqrt{n}(1-\lambda_n)\delta_n^2 \left[\frac{2}{9}(1-\lambda_n+\lambda_n^2)\sqrt{n}(1-\lambda_n)\delta_n^4 - M_2\right] && (111) \\
&\geq \sqrt{n}(1-\lambda_n)\delta_n^2 \left(\frac{1}{6}\sqrt{n}(1-\lambda_n)\delta_n^4 - M_2\right) && (112) \\
&\geq \sqrt{n}(1-\lambda_n)\delta_n^4 \frac{M'}{\bar{\delta}^2} && (113) \\
&\geq \frac{\sqrt{n}(1-\lambda_n)\delta_n^4}{6} \geq M' + M_2 > M', && (114)
\end{aligned}$$

where (111) uses the definition of B_n, (112) uses $1-\lambda_n+\lambda_n^2 \geq \frac{3}{4}$, (113) combines the definition of A_n with $\delta_n^2 \leq \bar{\delta}^2$, and (114) 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',

$$\Pr(n(1-\lambda_n)^2\delta^2\kappa^2 > M') \geq \frac{\epsilon}{2} \text{ i.o.}$$

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 8 *Assume the data is iid, $L_n(\theta)$ is continuous at $\forall\theta \in \Theta$ with probability 1, 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 $\bar{\sigma}^2 = \frac{\exp(\bar{\varkappa})}{\underline{\lambda}} = 2\exp(\bar{\varkappa})$ be an upper bound for $\max(\sigma_1^{*2}, \sigma_2^{*2})$, $\underline{\sigma}^2 = e^{-2\bar{\varkappa}}/(1+\frac{1}{4}\bar{\delta}^2)$ 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

$$\begin{aligned}
l(\theta) &= \log\left\{\lambda\frac{1}{\sqrt{\sigma_1^{*2}}}\exp\left[-\frac{(x-\mu_1^*)^2}{2\sigma_1^{*2}}\right] + (1-\lambda)\frac{1}{\sqrt{\sigma_2^{*2}}}\exp\left[-\frac{(x-\mu_2^*)^2}{2\sigma_2^{*2}}\right]\right\} \\
&\geq \lambda\log\left\{\frac{1}{\sqrt{\sigma_1^{*2}}}\exp\left[-\frac{(x-\mu_1^*)^2}{2\sigma_1^{*2}}\right]\right\} + (1-\lambda)\log\left\{\frac{1}{\sqrt{\sigma_2^{*2}}}\exp\left[-\frac{(x-\mu_2^*)^2}{2\sigma_2^{*2}}\right]\right\} \\
&\geq -\frac{1}{2}\log(\bar{\sigma}^2) - \frac{\lambda(x-\mu_1^*)^2 + (1-\lambda)(x-\mu_2^*)^2}{2\underline{\sigma}^2} \\
&\geq -\frac{1}{2}\log(\bar{\sigma}^2) - \frac{(|x|+\bar{\mu})^2}{2\underline{\sigma}^2},
\end{aligned}$$

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 $\bar{\mu}$. As a consequence,

$$\begin{aligned}
l(\theta) &= \log\left\{\lambda\frac{1}{\sqrt{\sigma_1^{*2}}}\exp\left[-\frac{(x-\mu_1^*)^2}{2\sigma_1^{*2}}\right] + (1-\lambda)\frac{1}{\sqrt{\sigma_2^{*2}}}\exp\left[-\frac{(x-\mu_2^*)^2}{2\sigma_2^{*2}}\right]\right\} \\
&\leq \log\left[\lambda\frac{1}{\sqrt{\sigma_1^{*2}}} + (1-\lambda)\frac{1}{\sqrt{\sigma_2^{*2}}}\right] = \log\left(\frac{1}{\sqrt{\underline{\sigma}^2}}\right).
\end{aligned}$$

Next, letting

$$d(x) = \frac{(|x|+\bar{\mu})^2}{2\underline{\sigma}^2} + \left|\log(\bar{\sigma^2})\right| + \left|\log\left(\frac{1}{\sqrt{\underline{\sigma}^2}}\right)\right|,$$

it is straightforward to see that $|l(\theta)| \leq d(x)$ and $E[|d(x)|] < \infty$. 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 9 *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

$$LR_n(\theta) = 2\frac{H_{3,n}}{\sqrt{n}}w_1 + 2\frac{H_{4,n}}{\sqrt{n}}w_2 - V_3 w_1^2 - V_4 w_2^2 + r_n(\theta),$$

so that

$$\begin{aligned}\frac{LR_n(\theta)}{w_1^2} &= 2\frac{H_{3,n}}{\sqrt{n}}\frac{1}{w_1} + 2\frac{H_{4,n}}{\sqrt{n}}\frac{w_2}{w_1^2} - V_3 - V_4\frac{w_2^2}{w_1^2} + \frac{r_n(\delta,\kappa,\lambda)}{w_1^2} \\ &\le 2\xi + 2\xi\frac{w_2}{w_1} - V_3 + \xi \\ &\le 5\xi - V_3 \\ &< 0\end{aligned}$$

because $V_3 = E[h_3^2] = 6$, which proves the result. □

Lemma 10 *Donsker property*

(10.1) $\sqrt{n}\left(\frac{1}{n}\tau^{-1}\frac{\partial^2 L(\delta(\tau,m),\kappa(\tau,m),1)}{\partial\lambda^2} - E\left[\tau^{-1}\frac{\partial^2 l(\delta(\tau,m),\kappa(\tau,m),1)}{\partial\lambda^2}\right]\right) = O_{p,(\tau,m)}(1)$.

(10.2) $\sqrt{n}\left(\frac{1}{n}\tau^{-1}\frac{\partial^3 L(\delta(\tau,m),\kappa(\tau,m),\lambda(\eta,\tau,m))}{\partial\lambda^3} - E\left[\tau^{-1}\frac{\partial^3 l(\delta(\tau,m),\kappa(\tau,m),\lambda(\eta,\tau,m))}{\partial\lambda^3}\right]\right) = O_{p,(\tau,m)}(1)$.

(10.3) $\frac{1}{\sqrt{n}}\frac{\partial^6 L_n(\delta,\kappa,1)}{\partial\lambda\partial\delta^i\partial\kappa^j} \Rightarrow \mathcal{G}^{[i,j]}(\delta,\kappa)$ *for* $i+j=5$.

(10.4) $\frac{1}{n}\tau^{-1}\frac{\partial^4 L(\delta(\tau,m),\kappa(\tau,m),\lambda(\eta,\tau,m))}{\partial\lambda^4} = O_{p,(\tau,m)}(1)$.

(10.5) $\tau^{-2}E\left[\frac{\partial^3 l(\delta(\tau,m),\kappa(\tau,m),\lambda(\eta,\tau,m))}{\partial\lambda^3}\right] = O_{(\tau,m)}(1)$.

(10.6) With μ *and* σ^2, $\frac{1}{\sqrt{n}}\frac{\partial^2 L(\rho_{n0})}{\partial\lambda\partial\mu} = O_p(1)$ *and* $\frac{1}{\sqrt{n}}\frac{\partial^2 L(\rho_{n0})}{\partial\lambda\partial\sigma^2} = O_p(1)$.

(10.7) 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 proof of (10.1) and (10.2) is similar to the proof of Proposition 1. Therefore, we only give the Taylor expansion of $\frac{\partial^2 l(\delta,\kappa,1)}{\partial\lambda^2}$ and $\frac{\partial^3 l(\delta,\kappa,1)}{\partial\lambda^3}$ to justify the normalization τ^{-1}, but omit the detailed steps. Specifically, fifth-order Taylor expansions yield

$$\begin{aligned}\frac{\partial^2 l(\delta,\kappa,1)}{\partial\lambda^2} =& h^4\left(\frac{1}{9}\delta^4 - \frac{1}{4}\kappa^2\right) + h_3\delta\kappa \\ &+ \sum_{i=3}^{4}\frac{1}{i!}\frac{\partial^{2+i}l(\delta,\kappa,1)}{\partial\lambda^2\partial\delta^i}\delta^i + \sum_{i+j=3,i\ge1,j\ge1}^{4}\frac{1}{i!j!}\frac{\partial^{2+i+j}l(\delta,\kappa,1)}{\partial\lambda^2\partial\delta^i\partial\kappa^j}\delta^i\kappa^j \\ &+ \sum_{i+j=5}\frac{1}{i!j!}\frac{\partial^{2+i+j}l(\tilde{\delta},\tilde{\kappa},1)}{\partial\lambda^2\partial\delta^i\partial\kappa^j}\delta^i\kappa^j,\end{aligned}$$

$$\frac{\partial^3 l(\delta,\kappa,1)}{\partial\lambda^3} = 8h^4\delta^4 + \sum_{i=3}^{4}\frac{1}{i!}\frac{\partial^{3+i} l(\delta,\kappa,1)}{\partial\lambda^3\partial\delta^i}\delta^i + \sum_{i+j=3, i\geq 1, j\geq 1}^{4}\frac{1}{i!j!}\frac{\partial^{3+i+j} l(\delta,\kappa,1)}{\partial\lambda^3\partial\delta^i\partial\kappa^j}\delta^i\kappa^j$$
$$+ \sum_{i+j=5}\frac{1}{i!j!}\frac{\partial^{3+i+j} l(\tilde{\delta},\tilde{\kappa},1)}{\partial\lambda^3\partial\delta^i\partial\kappa^j}\delta^i\kappa^j.$$

The proof of (10.3) is similar but much simpler, as it is not normalized by τ. To prove (10.4), it suffices to apply the uniform law of large numbers (see Lemma 2.4 of Newey and McFadden (1994)) and use

$$g(\tau,m) = \begin{cases} \tau^{-1}\frac{\partial^4 l(\delta(\tau,m),\kappa(\tau,m),\lambda(\eta,\tau,m))}{\partial\lambda^4} & \text{if } \tau\neq 0, \\ \lim_{\tau\to 0}\tau^{-1}\frac{\partial^4 l(\delta(\tau,m),\kappa(\tau,m),\lambda(\eta,\tau,m))}{\partial\lambda^4} = 24h^4 & \text{if } \tau = 0. \end{cases}$$

To see (10.5)

$$E\left[\frac{\partial^3 l}{\partial\lambda^3}\right] = -8960\delta^8 - 54\kappa^4 - 36\delta^2\kappa^2 + o(\tau^2).$$

As for (10.7), we can also show that evaluated at $\tilde{\rho}$

$$\frac{1}{n}\frac{\partial^3 L_n}{\partial\lambda^2\partial\mu} = -\frac{32}{3\hat{\sigma}}\delta^4\hat{H}_3 + \frac{2}{\hat{\sigma}}\kappa^2\hat{H}_3 + o_p(\tau),$$

$$\frac{1}{n}\frac{\partial^3 L_n}{\partial\lambda^2\partial\sigma^2} = -\frac{16}{3\sigma^2}\frac{1}{n}\hat{H}_4\delta^4 + \frac{1}{\sigma^2}\frac{1}{n}\hat{H}_4\kappa^2 - \frac{3}{2\sigma^2}\frac{1}{n}\hat{H}_3 + o_p(\tau),$$

where

$$\hat{H}_3 = \sum_i \hat{y}_i(\hat{y}_i^2 - 3),$$
$$\hat{H}_4 = \sum_i \hat{y}_i^4 - 6\hat{y}_i^2 + 3,$$
$$\hat{y}_i = \sum_i \frac{y_i - \hat{\mu}}{\hat{\sigma}},$$

whence we prove the desired result. □

Lemma 11 *$\left|\frac{1}{36}\delta^4 - \frac{1}{8}\kappa^2\right| \to 0$ 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. (2020): "Hypothesis tests with a repeatedly singular information matrix", CEMFI Working Paper 2002.

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. (1993): "Tests for parameter instability and structural change with unknown change point", *Econometrica* 61, 821–856.

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.

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.

Cameron, S.V. and Heckman, J.J. (1998): "Life cycle schooling and dynamic selection bias: models and evidence for five cohorts", *Journal of Political Economy* 106, 262–311.

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, 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.

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.

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.

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.

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.

Rotnitzky, A., Cox, D.R., Bottai, M. and Robins, J. (2000): "Likelihood-based inference with singular information matrix", *Bernoulli* 6, 243–284.

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. Sample size $n = 125$.

		Higher-order moments		Testing procedures: Score-type tests							Likelihood ratio test			Other tests	
				Θ_1		Θ_2			Θ_3						
δ	$\varkappa$	$E(x^3)$	$E(x^4)$	LM_1	JB	LM_2	$LM_{a,2}$	$LM_{b,2}$	LM_3	$LM_{a,3}$	Θ_1	Θ_2	Θ_3	KS	ACS
								λ =.75							
0	-2	0.00	3.68	31.0	28.6	19.1	11.3	19.2	46.1	35.9	11.6	11.6	42.7	27.2	35.1
0	.5	0.00	3.18	10.2	10.3	10.4	7.5	10.5	11.0	11.3	6.2	5.8	11.1	5.5	7.1
2	-2	0.21	2.29	42.9	26.1	23.0	60.5	10.2	0.2	0.1	87.1	72.8	0.1	63.2	62.7
2	0	-0.32	2.88	23.7	17.8	19.7	32.0	19.2	2.6	2.5	35.9	39.5	2.5	27.5	37.6
2	.5	-0.60	3.34	70.3	65.9	66.0	77.4	65.6	19.0	18.5	75.8	78.5	19.6	59.9	78.7
								λ =.95							
0	-2	0.00	3.12	6.8	7.1	7.0	5.6	7.0	8.8	8.1	4.2	4.6	8.7	5.8	6.7
0	.5	0.00	3.06	6.4	6.4	6.9	5.7	6.9	6.7	6.6	5.3	5.1	6.6	4.8	5.3
2	-2	-0.07	2.81	3.9	2.1	2.8	4.7	2.7	1.6	1.5	8.1	9.0	1.4	6.6	7.9
2	0	-0.26	3.38	23.2	24.7	23.1	22.4	23.1	19.5	20.0	16.4	16.0	20.0	10.2	17.4
2	.5	-0.40	3.88	42.6	44.4	42.8	39.3	42.8	39.0	39.7	31.4	28.7	39.1	14.9	29.2

Notes: Results based on 10,000 replications with critical values computed using 100,000 replications. 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. 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: Bootstrap-based size corrected rejection rates at 5% significance levels for Gaussian null hypothesis. Sample size $n = 500$.

		Higher-order moments		Testing procedures											
				Score-type tests							Likelihood ratio test			Other tests	
				Θ_1		Θ_2			Θ_3						
δ	$\varkappa$	$E(x^3)$	$E(x^4)$	LM_1	JB	LM_2	$LM_{a,2}$	$LM_{b,2}$	LM_3	$LM_{a,3}$	Θ_1	Θ_2	Θ_3	KS	ACS
							$\lambda = .75$								
0	-2	0.00	3.68	82.8	70.9	34.5	10.4	35.1	93.7	82.7	54.3	34.4	91.9	84.5	89.9
0	.5	0.00	3.18	15.5	16.0	14.2	6.6	14.5	19.2	19.6	7.7	6.8	19.4	5.8	8.7
2	-2	0.21	2.29	100.0	99.7	99.3	99.9	41.5	0.0	0.0	100.0	100.0	0.0	100.0	100.0
2	0	-0.32	2.88	86.7	83.0	79.2	91.0	74.2	1.3	1.2	93.3	94.1	1.2	81.4	94.3
2	.5	-0.60	3.34	100.0	100.0	99.9	100.0	99.9	38.8	38.8	99.9	100.0	42.5	99.6	100.0
							$\lambda = .95$								
0	-2	0.00	3.12	9.3	9.6	7.5	4.7	7.7	13.7	12.2	5.1	5.0	14.0	8.5	9.4
0	.5	0.00	3.06	8.1	7.9	8.3	5.3	8.5	9.1	9.1	5.3	5.1	8.7	5.3	5.8
2	-2	-0.07	2.81	16.3	6.6	5.0	14.9	3.1	0.6	0.4	30.6	21.6	0.4	15.5	22.2
2	0	-0.26	3.38	59.2	63.3	57.9	57.6	58.4	41.9	44.0	53.6	57.0	44.2	28.9	49.6
2	.5	-0.40	3.88	87.6	89.2	87.2	83.6	87.5	79.0	80.7	82.8	84.2	80.1	45.5	73.0

Notes: Results based on 10,000 replications with critical values computed using 100,000 replications. 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. KS denotes the Kolmogorov-Smirnov test and ACS the CGMM test proposed in Amengual, Carrasco and Sentana (2020) with Tikhonov regularization parameter $\alpha = .01$ and scale parameter $\omega^2 = 1$.

Table 3: Application to Mincer equations

Specification		(1)		(2)		(3)	
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
KS		0.2	.66	0.4	.36	0.5	.05
ACS		-0.6	.19	-0.9	.48	-0.5	.13

Notes: CPS85 dataset provided by the Berndt (1981). (1) refers to women and men, (2) refers to men only, and (3) women only. For both, the score-type tests and the likelihood ratio test, the three different parameter spaces are

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

JB skew (JB kurt) refers to the Jarque-Bera skewness (kurtosis) component of the Jarque-Bera (1980) test. 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 $\omega^2 = 1$. LM's and LR's are defined in Section 3.